BIOLOGICAL INVESTIGATION AND PREDICTIVE MODELLING OF FOAMING IN ANAEROBIC DIGESTER

IFEYINWA RITA KANU

Submitted for the degree of Doctor of Philosophy

Heriot Watt University

School of Energy, Geoscience, Infrastructure and Society (EGIS)

May 2018

Abstract

Anaerobic digestion (AD) of waste has been identified as a leading technology for greener renewable energy generation as an alternative to fossil fuel. AD will reduce waste through biochemical processes, converting it to biogas which could be used as a source of renewable energy and the residue bio-solids utilised in enriching the soil. A problem with AD though is with its foaming and the associated biogas loss. Tackling this problem effectively requires identifying and effectively controlling factors that trigger and promote foaming.

In this research, laboratory experiments were initially carried out to differentiate foaming causal and exacerbating factors. Then the impact of the identified causal factors (organic loading rate-OLR and volatile fatty acid-VFA) on foaming occurrence were monitored and recorded. Further analysis of foaming and non-foaming sludge samples by metabolomics techniques confirmed that the OLR and VFA are the prime causes of foaming occurrence in AD. In addition, the metagenomics analysis showed that the phylum bacteroidetes and proteobacteria were found to be predominant with a higher relative abundance of 30% and 29% respectively while the phylum actinobacteria representing the most prominent filamentous foam causing bacteria such as *Norcadia amarae* and *Microthrix Parvicella* had a very low and consistent relative abundance of 0.9% indicating that the foaming occurrence in the AD studied was not triggered by the presence of filamentous bacteria.

Consequently, data driven models to predict foam formation were developed based on experimental data with inputs (OLR and VFA in the feed) and output (foaming occurrence). The models were extensively validated and assessed based on the mean squared error (MSE), root mean squared error (RMSE), R^2 and mean absolute error (MAE). Levenberg Marquadt neural network model proved to be the best model for foaming prediction in AD, with RMSE = 5.49, MSE = 30.19 and R^2 = 0.9435. The significance of this study is the development of a parsimonious and effective modelling tool that enable AD operators to proactively avert foaming occurrence, as the two model input variables (OLR and VFA) can be easily adjustable through simple programmable logic controller.

Dedication

I dedicate this thesis to almighty God, the source of all wisdom and knowledge.

Acknowledegement

My profound gratitude goes to my supervisor, Professor Adebayo J. Adeloye whose ideas, constructive criticism and benevolent appreciation has enabled me to bring out the best from my research work.

I thank Dr Thomas Aspray for accepting that I should use his laboratory for my experiments. I am very grateful for all his support and understanding in the early days of getting acquainted to the laboratory practise. I am grateful to Evans Lawrence and Victor Ude for their guidance and support.

Thanks to Professor Scott Arthur and Dr Rabee Rustum for reviewing my thesis and offering feedback, your comments have brought in different perspective that has enriched my thesis.

To the amazing staff at Veolia Water: Paul Banfield, Michael Skroblin and Jenniffer Lawson (nee:LittleJohn) I am so grateful to all of you. My PhD would not have been a success without your support all the way from historical data to sample collections and field trips.

I wish to say thanks to Tom Ferguson and the technicians: Sorley Graham and Julian Ramos for always making out time, vehicle, and safety equipment to pick samples from the wastewater treatment plants for my experiments.

To all my colleagues whom we have embarked on the PhD journey, it has been a great experience working with diverse students from various background and culture; a gift that Heriot Watt University offers to every student of which I am very grateful.

Most importantly, my PhD journey will not be complete without my awesome family who had stood by me through the thick and thin. Thanks to the best mums in the world Mrs Mary-Regina Ifeanyi and Lady Comfort Kanu for all their motherly support. To my husband Chika Osita Kanu and our lovely children: Ifunanyachi, Chikamso, Chikamma, and Chikatara thanks a lot for understanding with me through the moments that I had to work so hard to finish my thesis, you have been the best!

ACADEMIC REGISTRY

Research Thesis Submission

Name*:*	Ifeyinwa Rita Kanu		
School:	Energy Geoscience Infrastructure and Society		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	Doctor of Philosophy

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	07/05/2018

Submission

Submitted By *(name in capitals):*	Ifeyinwa Rita Kanu
Signature of Individual Submitting:	
Date Submitted:	07/05/2018

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission (Handed in to SSC; posted through internal/external mail):			
*E-thesis Submitted (**mandatory for final theses**)*			
Signature:		Date:	

Table of Contents

Figures

Tables

List of Acronyms

Acronyms	Meaning
ABS	Absorbance
AD	Anaerobic Digester
ADM	Anaerobic Digestion Model
ADV	Additional Digester Volume covered with foam
AF	Allenfern Wastewater Treatment Plant
AI	Artificial Intelligence
ALK	Alkalinity
AM	Aceticlastic Methanogens
ANN	Artificial Neural Network
AR	Auto-Regressive Model
ARIMA	Auto-Regressive Integrated Moving Average
ARMA	Auto-Regressive Moving Average
AS	Activated Sludge
BMP	Biochemical Methane Potential
BOD	Biological Oxygen Demand
BPR	Biogas Production Rate
BSI	British Standard Institution Publicly Available Specification
BVS	Biodegradable Volatile Solid
C	Concentration of VS in digester feed
CE	Capillary Electrophoresis
CMC	Critical Micelle Concentration
COD	Chemical Oxygen Demand
DEFRA	Department for Environment, Food and Rural Affairs
DNA	Deoxyribonucleic Acid
EPS	Extra Polymeric Substances
F	Foaming reactors 2 hours after feed
Fe	Feed sample
FFNNs	Feed-forward Neural Network
FI	Filamentous Index
FIS	Fuzzy Inference System
FISH	Fluorescent In-Situ Hybridisation
FL	Fuzzy Logic
FLM	Fuzzy Logic Model
FOG	Fat, Oil and Grease Waste
FP	Foaming Potential
G	Gaussian Membership Function
GB	Gbell Membership Function
GC	Gas Chromatography
GHG	Green House Gas
GWH	Giga Watt Hour

Acronyms	Meaning
HA	Hatton Wastewater Treatment Plant
HAc	Acetic acid
HM	Hydrogenotrophic Methanogens
HPLC	High Pressure Liquid Chromatography
HRT	Hydraulic Retention Time
I	Inoculum
IC	Ion Chromatography
IWA	International Water Association
KSOM	Khonen Self-Organising Map
LC	Liquid Chromatography
LCFA	Long Chain Fatty Acid
MAE	Mean Absolute Error
MF	Membership Function
MFLM	Mamdani Fuzzy Logic Model
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
MSE	Mean Squared Error
MVF	Maximum Volume of Foam observed in the reactor
NB	Newbridge Wastewater Treatment Plant
NBD	New Bridge Digestate
NBP	New Bridge Pre-digester feed
NF	Non-Foaming Reactor 2 hours after feed
NGS	Next Generation Sequencing
NMR	Nuclear Magnetic Resonance
NN	Neural Networks
NNBR	Neural Network Bayessian Regularisation
NNFC	National Non-Food Crops Centre
NNFIS	Neural Network Fuzzy Inference System
NNLM	Neural Network Levenberg Marqudart
NNSC	Neural Network Scaled Conjugate
OLR	Organic Loading rate
Omics	Integrated Omics
OTU	Operational Taxonomic Unit
PABR	Periodic Anaerobic Baffled Reactor
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCS	Principal Component Scores
PF	Prior to feeding Foaming reactors
pH	Hydrogen ion concentration
PID	Proportional Integral Derivative
pKA	Acid Ionisation Constant
PLC	Programmable Logic Controller

Acronyms	Meaning
PNF	Prior to feeding Non-Foaming reactors
PS	Primary Sludge
Q	Influent flow rate
QIIME	Quantitative Insight into Microbial Ecology
r	Correlation Coefficient
RMSE	Root Mean Squared Error
RNA	Ribonucleic Acid
RNN	Recurrent Neural Network
SF	Seafield Wastewater Treatment Plant
SFD	Seafield Digestate
SFP	Seafield Pre-digester feed
SP	Soluble Protein
SRT	Solid Retention Time
STP	Standard Temperature and Pressure
TDS	Total Dry Solid Concentration
TEM	Transmission Electron Microscopy
TKN	Total Kajelda Nitrogen
Tp	Trapezoidal Membership Function
Tr	Triangular Membership Function
TS	Total Solid Concentration
tVFA	Total Volatile Fatty Acid oncentration
TWAS	Thickened Waste Activated Sludge
UPLC	Ultra-Performance Liquid Chromatography
V	Bioreactor volume
VA	Volatile Acid Concentration
VFA	Volatile Fatty Acid Concentration
VS	Volatile solid Concentration
VSR	Volatile Solid Reduction
WAS	Waste Activated Sludge
WRAP	Waste Resource Action Plan
WWTP	Wastewater Treatment Plant

List of Publications

Journal Articles

1. Kanu, I., Aspray, T. & A., A., 2015a. Understanding and predicting foam in anaerobic digester. *International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering,* 9(10), pp. 1056-1060.

Conferences

1. Kanu, I., Aspray, T. J., Scott, A. & Adeloye, A., 2015. *Laboratory experiment to select variables for predicting foaming in anaerobic digester.* Edinburgh, In proceedings of the Infrastructure and Environment Scotland Postgraduate Conference, Heriot Watt university.

2. Towards a new predictive model of foaming in anaerobic digesters; AQUA ENVIRO 20th European Biosolid conference 2015

3. Laboratory experiment to select variables for predicting foaming in anaerobic digester; Infrastructure and Environment Scotland 3rd Postgraduate Conference 2015

4. A further insight to foaming in anaerobic digester; ICEEWM 17th International conference on environment, and waste management 2015.

Newspaper Articles

1. Food-waste-could-cost-us-the-earth; Scotsman, 29 August 2017 http://www.scotsman.com/news/opinion/ifeyinwa-kanu-food-waste-could-cost-us-the-earth-1-4544460

Patent

1. Title: Anaerobic Digester; Application Number: GB1705768.8, Country: UK; Applicant: Ifeyinwa Rita Kanu; Date: 10 April 2017

Chapter 1 : Introduction

1.1 Problem Background

Wastewater treatment aims at removal of significant quantity of major pollutants particularly oxygen demanding substances such as the biochemical oxygen demand (BOD), prior to discharge of the effluent into the environment (Rustum & Adeloye, 2009). This is achieved through a series of physical, chemical, and biological treatment processes as wastewater undergoes preliminary, primary, secondary and tertiary stages of treatment. Consequent to wastewater treatment is an effluent which is safe to be discharged into the environment or re-used but there is also a large volume of sludge that requires further treatment before disposal to the environment.

Sludge production during wastewater treatment process occurs mainly during primary treatment and secondary treatment; thus the associated sludges are usually known as primary sludge (PS) and waste activated sludge (WAS), respectively. An activated sludge (AS) wastewater treatment plant in Scotland with a population equivalent of 806,883 and inflow rate of 335,226m^3/d generates sludge in the range of 640m^3/d and 360m^3/d of 5% thickened PS and dewatered WAS respectively (Sourced data_Seafield performance log).

In Scotland, water and wastewater treatment consumes about 445 GWHof electricity per annum accounting for 76% of the total greenhouse gas (GHG) emission of Scottish Water. Forty-five percent of this relates to wastewater treatment (Scottish Water, 2016). On the contrary, the energy in sewage (thermal, chemical and mechanical) is 2-4 times the amount of energy employed in treating it (Banfield & Littlejohn, 2013) thus there is need to efficiently harness this energy to enable wastewater treatment works become energy neutral/surplus.

The Environment Agency, the Scottish Environment Protection Agency and the Environment and Heritage Service of the Department of the Environment for Northern Ireland are the environmental regulators of the water industry for the UK. Hence, they regulate discharges from wastewater treatment works and combined sewer overflows, which cannot be made without conditional ‘discharge consent’ or in Northern Ireland ‘registered standards’ being issued. These authorisations and conditions are set to minimise the adverse

effects of pollution on the receiving waters, and meet standards of European Directives. The requirement of the Directive in relation to discharges from urban wastewater treatment plants to sensitive areas subject to eutrophication are as shown in table 1-1.

Table 1-1: European Commission Directive on Urban Wastewater treatment 1998 (Amendment to Directive 91/271/EEC)

Parameters	**Concentration**	**Minimum percentage of reduction**	**Reference method of measurement**
Total phosphorus	2 mg/l (10,000-100,000 p.e.)	80	Molecular absorption spectrophotometry
	1 mg/l (>100,000 p.e.)		
Total nitrogen	15 mg/l (10,000 to 100, 000 p.e.)	70-80	Molecular absorption spectrophotometry
	10 mg/l (> 100,000 p.e.)		

The result is much cleaner water being returned to the environment, and sludge that contains the organic matter and dead bacteria from the treatment process. DEFRA (2002) report highlighted that there was a 13.27% increment in sludge tonnage generated in UK from a total of 997, 673 in 1998 to 1,130,066 in 2002. Thus it was concluded in the report that higher standards of treatment generate more sludge, hence, we can expect the amount of sludge generated from wastewater treatment process to continue rising because of stringent requirements and standards.

Historically in the UK, about a quarter of sludge was either dumped at sea or discharged to surface waters. This was banned from 1998 under the Urban Wastewater Directive because it was considered environmentally unacceptable. The quickest response to use of sludge following this directive was a 52% on farmland, 21% on incineration, 17% on landfill and the rest of 10% applied for various purpose (DEFRA, 2002). With the challenges of using sludge on farmland due to contamination British Standard Institution Publicly Available Specification (BSI PAS) 100 for certification of waste material treated to be used in farm or for growing crops was developed. This was launched in November 2002 and was developed jointly by Waste Resource Action Plan (WRAP) and The Association for Organics Recycling.

On the other hand, there is the European Commission Directive on the Landfill of Waste (1999/31). The main goal of this Directive on the landfill of waste was to encourage waste prevention by its recycling and recovery. Hence, the Directive establishes high standards for the treatment of waste in the EU and discourages, as much as it can, their disposal which is a source of water, air, and ground pollution with harmful effects for human health. One of the essential elements is a gradual reduction in municipal biodegradable waste. This waste will have to be reduced to 75% of its weight by 2002 compared to 1993 and banned by 2020. Sludge generated from the treatment of Municipal sewage sludge falls under this category thus can neither be landfilled nor incinerated.

These regulations have resulted to need for further treatment of sludge. This could be achieved through various sludge stabilisation processes (SSP). Sludge stabilisation process such as; alkaline stabilisation, anaerobic digestion, aerobic digestion, authothermal thermophilic digestion and composting (Eddy/AECOM, 2014). Of these, anaerobic digestion has been the most commonly used because of the added advantage of generating heat or electricity (from biogas) as well as useful biosolids (stabilised sludge).

Biogas which is rich in methane is a clean source of fuel when compared to fossil fuels and its use for energy production is thus certain to result in reduction of pollution and environmental degradation (Khalid, et al., 2011).

The biosolids have also found applications in enhancing the fertility of agricultural soils due to the presence of core plant nutrients such as Nitrogen (N), Phosphorus (P), Potassium (K) and Magnesium (Mg) (Leven, et al., 2012). Furthermore, amongst all the renewable energy sources such as solar and wind, biogas has been found to be the most environmentally friendly by reducing the impact of waste on the environment as well as generating clean energy. Therefore, there is need to intensify efforts in ensuring that the process becomes more efficient and it is adequately optimised.

AD traditionally evolved for the treatment of sludge resulting from wastewater treatment operations. Although wastewater treatment AD still accounts for majority of AD operations in the UK, there have been diverse application of AD in other industries including animal manures, food waste and energy crops as demand for cleaner energy soars (Li et al., 2015a; Moset et al., 2015; Weinrich & Nelles, 2015). In UK for example, as of September 2011 to March 2017, the number of anaerobic digester (AD) in operation rose from 214 to 392, a rise of almost 45% as in figure 1-1.

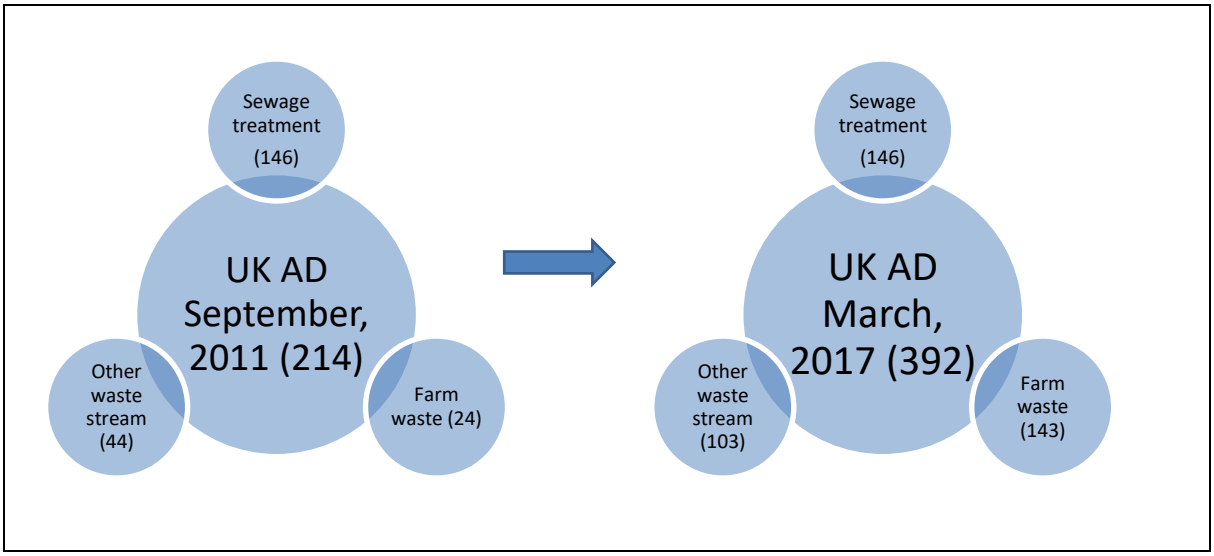


Figure 1-1: Anaerobic digester growth in UK

The 146 AD for the wastewater treatment processes up to 1.1 Million tonnes of feedstock generating up to 110MW of electricity per annum (WRAP & NNFCC, 2015).

However, foaming has been a continuous problem amid the various challenges experienced in the operation of AD. This was illustrated in a survey by the American Society of Civil Engineers reporting half of the AD to have experienced foaming once during their operating lives (Filbert, 1985). A survey of foaming in ADs in wastewater treatment plants in USA carried out from April to August 2011 showed that out of the 39 plants surveyed, 32 had experienced foaming in the past five years or were presently undergoing foaming (Subramanian, et al., 2012). The identified causes include: presence of foam causing filament, fats/oil/grease (FOG) and feed sludge quality (Subramanian et al, 2012).

A similar survey was carried out in Spain, in which out of 38 plants that responded to the survey, 23 of them had experienced foaming with the causes being attributed to sludge characteristics and poor operating factors such as variation in feed quality/quantity, temperature, pH etc. (Rodríguez-Roda et al., 2013). Nevertheless, neither Subramanian et al. (2012) nor Rodriguez-Roda et al. (2013) were able to establish a specific relationship between process parameters and causes of foaming.

The complications encountered in solving the problem of foaming in AD stem from the fact that it is occurring in an environment of microbiological complexity that is immensely influenced by variation in operational conditions. This has resulted to lack of sufficient knowledge on efficient remedial actions for mitigating foaming in anaerobic digesters (Subramanian et al, 2012). On this premise, modelling becomes a useful tool as controlling treatment plants through modelling is technically the most feasible and probably the least costly way of attaining a sustainable improvement in performance (Rustum & Adeloye, 2012, Dalmau, et al., 2009).

Currently full-scale AD uses Proportional Integral Derivative (PID) system in controlling the digestion process. Steyer et al., (1999) proposed a control strategy for highly loaded anaerobic digestion processes by analysing the system disturbance due to increased influent flow rate. Based on the analysis, the biogas output flow rate and pH controlled the control system functions (Steyer, et al., 1999). This control system was mainly flawed because the

variables that were selected as control (biogas and pH) are not reliable cursor as we have systematically established in this study

Dalmau et al., (2009) applied Mamdani Fuzzy logic to model the risk of foaming in AD using heuristic information based on the digester pressure as an indication of foaming occurrence while the inputs were OLR, variation in OLR and presence of filamentous microorganism. Again, this model was flawed as there was no clear data that represent foaming occurrence as this was inferred based on AD pressure records. In addition, input variables (OLR and variation in OLR) were dependent variables that should not be applied at the same time to develop a model.

Based on these fundamental researches, the following research gap has been identified in this study:

1. There is lack of clear understanding on the actual factors causing foaming in AD as well as how other factors contribute to the exacerbation and sustenance of the foaming occurrence.

2. Lack of comprehensive model of foam formation in AD since there was existence of only one AD model integrating foaming occurrence of which the variables and the source of data is not considered adequate for such purpose.

Seeking to address this research gap has led to the implementation of extensive laboratory experiment to acquire in-depth knowledge on microbial complexities existing in foaming and non- foaming reactors and the dynamics of foaming development under varying system operational parameters. The experiments were carefully designed by analysing existing data and carrying out initial experimentation to identify important design and operating variables that favour foaming. The data collected was used to calibrate, test, and validate predictive models of foam formation in AD.

1.2 Aim and Objectives

The aim of this project is to establish a suitable artificial intelligence (AI) model for predicting foam formation in AD by isolating design and operating

variables that favour foaming in AD and extensive knowledge of microbial population present in foaming and non-foaming AD. The specific objectives are to:

1. Extensively review the literature on AD operation and modelling to ascertain factors relating to AD foaming, limitations in existing models *to establish knowledge gaps that will assist in developing the research methodology*

2. Identify suitable full-scale AD in Scotland and collect data on operation parameters including foaming history and biogas production inventory. Carry out statistical analysis including correlation analysis *to identify statistically significant process and design variables affecting foam formation.*

3. Design and carry out laboratory experiment that assesses the variation in foam formation to variations in the identified significant process variables in order *to identify the main process parameters underpinning the foaming process in AD.*

4. Carry out metagenomics and metabolomics analysis of foaming and non-foaming samples *to enhance knowledge on microbial composition and their influence on foaming.*

5. Formulate, calibrate, and validate model for predicting the formation and extent of foaming in AD with a view *to improving the efficiency of biogas from such plants.*

6. Make recommendations.

1.3 Study focus and outline

This study is focussed on AD treating sewage sludge. An outline of the chapters covered in this study is as follows:

Chapter 2: Literature review. This unit started with an introduction to activated sludge (AS) wastewater treatment process and the important role of AD in treating the sludge generated during AS process.

The various classification and biochemical process involved in AD process were discussed with focus on foaming and its effect on the performance of AD.

This was followed by detailed study of 'omics' analysis and its application as a powerful tool to comprehensively reveal the microbial structure and metabolic pathway of complex AD sludge.

Review of experimental design, data collection/analysis and synthetic data generation were presented. Modelling AD mechanistically and empirically were also reviewed

Chapter 3: Methodology. The chapter describes the materials and method used in this research which include; sources of historical data and sludge samples (full scale foaming and non-foaming AD treating Sewage sludge), experimental design and statistical/ 'omics' analysis adopted, procedures for synthetic data generation and development of AI model using MATLAB programming language in addition to discussions on model performance evaluation.

Chapter 4: Selecting the most appropriate variable for foaming in AD. In this chapter, statistical analysis of historical data and preliminary laboratory experiment was carried out to identify factors related to AD foaming, distinguishing causal factors from those factors enhancing foaming occurrence and monitor foaming propensity by analysing samples collected from full scale foaming and non-foaming digesters.

Chapter 5: Data collection for modelling foaming in AD. Laboratory scale AD experiment designed to evolve a foaming and non-foaming digester was carried out and data collected. Statistical analyses were carried out to establish those factors that are crucial for modelling foaming in AD.

Chapter 6: Metagenomic and Metabolomics analysis. These 'omics' analysis was carried out using next generation high throughput sequencing to compare the result from the laboratory analysis in order to further establish the contribution of microbiome, metabolites and operating conditions on AD

foaming as well as validate the results generated from the laboratory experiment.

Chapter 7: Predictive modelling of foaming in AD. Using AI techniques such as Fuzzy logic (FL) and Neural Networks (NN), predictive model of foaming in AD were formulated, calibrated, and validated. The performances of the various AI models were evaluated and the most suitable model for predicting foaming in AD was selected.

Chapter 8: Conclusions and recommendations. In this chapter, discussions were carried out and recommendations for further research put forward.

Chapter 2 : Literature review

In this chapter, we reviewed the sources of feed to AD treating sewage sludge, the different anaerobic digesters with regards to structure and operating conditions and how they impact on the digester efficiency. The application of 'omics' study in achieving a better understanding of the digester condition was explored. Modelling tools were reviewed and search for existing AD models were carried out to establish the gap in AD modelling.

2.1 Wastewater treatment

Between 1830 and 1840, England was among the first countries to react to issues relating to wastewater. This was in response to Cholera outbreak as raw sewage dumped on the streets and stored in cess pits mixed with water supplies used for drinking. Sewers were constructed to dispose untreated foul sewage and surface water into rivers (Butler & Davies, 2011) whose self-purification capacities were inadequate to cope with the high levels of pollution in the discharges. With more stringent policies on waste handling and disposal, wastewater treatment process has since undergone series of developments and modifications as illustrated in figure 2-1 (Eddy/AECOM, 2014)

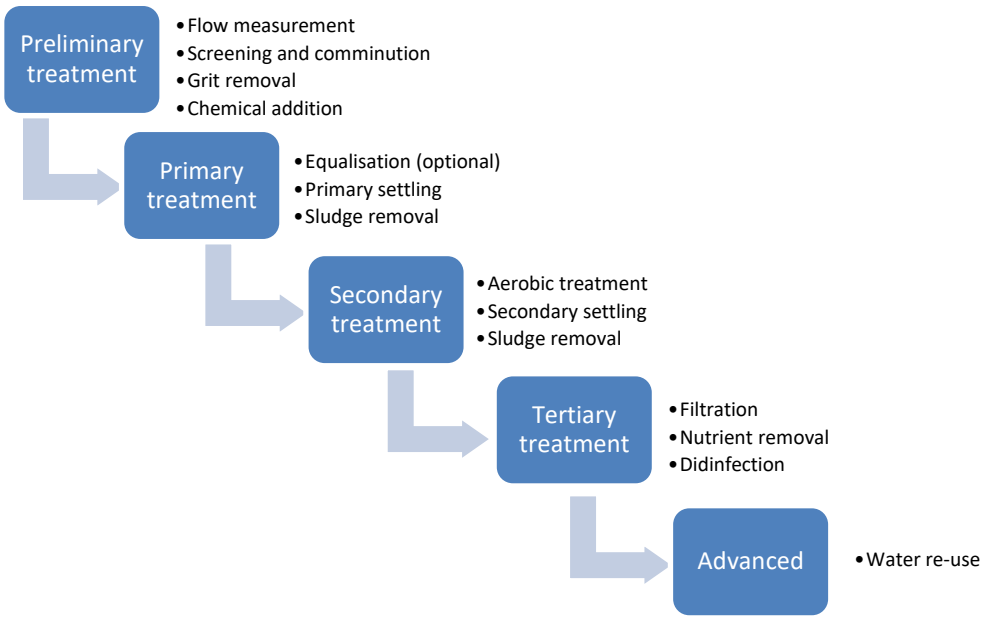


Figure 2-1: Stages in Activated Sludge wastewater treatment process

During preliminary treatment, removal of materials that could result in excessive wear to machineries occurs thereby saving valuable space within

the treatment plant and downstream processes. Flow equalisation is another aspect of preliminary treatment and can be achieved in wastewater treatment operations using storm tanks. This helps to prevent hydraulic overload of subsequent treatment which could result in loss of efficiency of the biological treatment stage. Having storm tanks also prevents the discharge of the first-foul-flush into the environment thereby improving quality of overflows discharged into the river (Rustum & Adeloye, 2012).

Since primary treatment is designed to remove from wastewater, settleable and floatable solids, its efficiency has a direct influence on the ensuing biological and sludge treatment. Using large basins under gravity settling, primary sludge (PS) is produced that comprises 50 to 70 % of the incoming total suspended solids (Eddy/AECOM, 2014). Generally, PS is usually grey and slimy with an extremely offensive odour (Eddy/AECOM, 2014).

The central drive for secondary treatment is to provide BOD removal beyond what is achievable by primary treatment (Eddy/AECOM, 2014) with nutrient (nitrogen and phosphorus) removal attainable in some setup (Pitman, 1991). During secondary treatment, microorganisms convert dissolved, suspended, and colloidal organic wastes to more stable solids known as secondary sludge. The microorganisms are nurtured by organic pollutants in the wastewater and can be: heterotrophic and autotrophic bacteria, yeasts, algae, fungi, filamentous bacteria, and protozoa (Eikelboom, 2000).

A balance in the population of these microorganisms is essential for the formation of good floc that will settle to stable solids- sludge. The quantity of filamentous bacteria present should suffice solely for the flocculent zoogleal microorganisms to attach and form a floc. Inadequate filamentous bacteria results in weak flocs while excessive filamentous bacteria which tend to extend from the floc into the medium prevent the flocs from coming together thereby resulting in poor compaction referred to as bulking sludge (Sezign, et al., 1978).

With regards to secondary treatment reactor set up, the microorganisms can grow either suspended in the wastewater e.g. activated sludge or fixed on the surface of a media e.g. trickling filter (Eddy/AECOM, 2014). Secondary

sludge is known as waste activated sludge (WAS) in suspended growth and humus in fixed film.

Comparing fixed film and suspended growth systems, greater volume of sludge is produced using suspended growth. Typically, WAS appears as a brown flocculant with an offensive odour if in good condition although it can go septic quickly (Eddy/AECOM, 2014).

Furthermore, PS compared with WAS shows enhanced degradability (Barber, 2014). This is attributable to the differences in their inherent characteristics which reflects their composition as shown in figure 2-2.

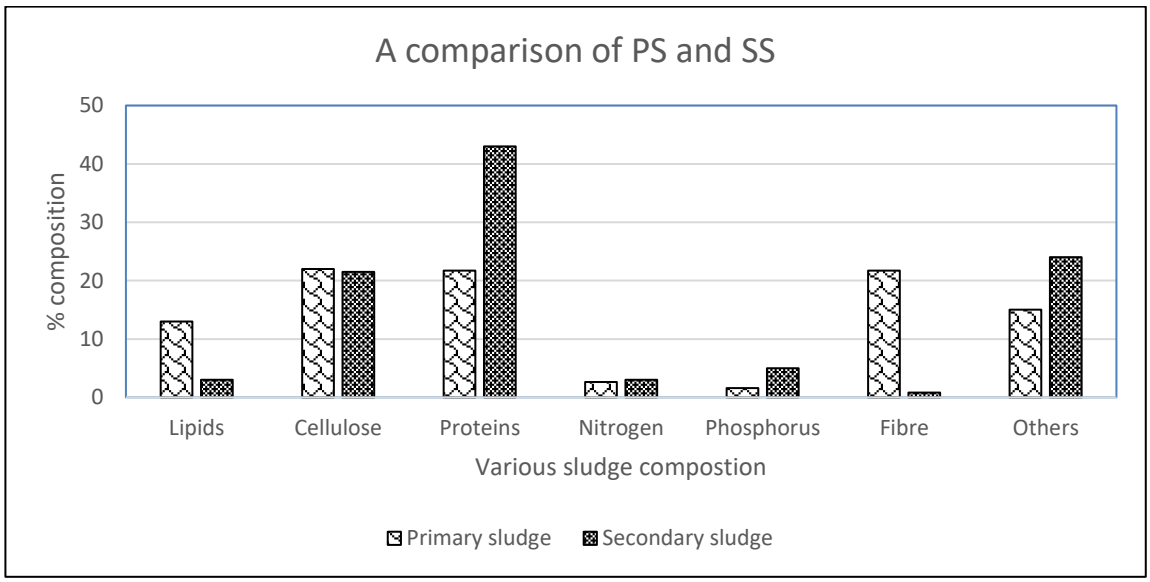


Figure 2-2: A comparison of % composition of PS and WAS adapted from Barber (2014)

It could be observed from figure 4 that the highest variation in primary and secondary sludge is in the protein at a difference of over 20%. This higher concentration of protein in secondary sludge and the fact that protein degradation is slower compared to carbohydrate and lipids further support the notion that WAS is less degradable (Barber, 2014).

Based on its composition, WAS has a greater concentration of bound water, volatile solids (VS), extra-cellular polymeric substances (EPS) and high molecular weight material (Barber, 2014). On the contrary, PS has larger particle size due to the presence of more inert materials and fibres as shown in figure 2-2. The resultant effect is the ability to achieve greater thickening

for PS compared to WAS which permit the use of gravity thickening for PS while WAS passes through dewatering by centrifuge belt so that volume of water contained in them could be reduced to about 5% TS prior to feeding them into the AD. The PS and WAS after thickening are sent to the anaerobic digester for further processing.

As seen in table 2-1, not only is the PS more amenable to low cost gravity thickening, its biogas yield also exceeds that of the WAS

Table 2-1: A comparison of response of PS and WAS main characteristics Barber (2014)

Primary sludge (PS)	Waste Activated Sludge (WAS)
Easier to thicken_ 25 to 35%	More difficult to thicken_14 to 18%
50-60% volatile solid reduction	30-45% volatile solid reduction
25,700 kJ/kg volatile material	21,800 kJ/kg volatile material
0.983 Nm^3/kg of biogas/volatile solid destroyed	0.788 Nm3/kg of biogas/volatile solid destroyed

2.2 Anaerobic digestion basics

Anaerobic digestion (AD) is a multi-stage biochemical process which uses microorganism to stabilise organic matter in an environment devoid of free or dissolved oxygen. This reduces the volume of waste while producing biogas made up of approximately 60-65% methane, 35-40% carbon dioxide (biogas) and other impurities such as hydrogen sulphide, moisture, and siloxanes (WEFTEC, 2007; Miler Herman, 2003). The carbon dioxide and other gasses are removed prior to storage of the biogas for subsequent use as a source of energy. The key stages in sludge handling for anaerobic digestion are represented figure 2-3.

Figure 2-3: Stages in the treatment of sewage sludge for anaerobic digestion

2.3 Anaerobic digester classification

Anaerobic digesters can be categorised based on structure or operational condition. Structurally, AD can either be cylindrical or oval. Typically, cylindrical digesters have cone-shaped floors with slope of 1:4 to 1:6 that facilitates collection and removal of heavy sludge and can be equipped with gas holder cover if it is built with concrete (Eddy/AECOM, 2014). Oval shaped digesters are constructed with relatively steeper slopes typically 1:1. Usually, oval shaped digesters have conical bottom and a domed top. They are more difficult and expensive to build but involve lower operating cost due to reduced potential for grit amassing because of their steeper slope.

There are three major types of AD based on the feed to the digesters. They are: low-solids wet anaerobic digesters, high-solids wet anaerobic digesters, and dry anaerobic digesters (Bartell, et al., 2015) based on the total solid content (TS) of the feedstock. Thus low-solids wet digesters, high-solids wet digesters and dry digesters are characterised by TS of 5%, 20% and above 25% respectively (Eddy/AECOM, 2014).

In terms of the operational conditions, ADs are mainly classified based on temperature, degree of mixing and stages of operation involved. Temperature classification includes psychrophilic, mesophilic, and thermophilic, operating at temperatures 10-20^{o}C, 30-40^{o}C and 55^{o}C or higher respectively. When temperature of AD steps down from mesophilic to psychrophilic range, the composition of the microbial population can be very similar. In such circumstance, large numbers of mesophilic microorganism remain active with

low metabolic rates. It could then be argued that psychrotolerant rather than psychrophilic microorganisms are involved in the AD process as the existence of psychrophilic organisms in the AD remains unclear (van Lier et al., 1997).

Classification based on mixing includes mixed and unmixed digesters. Unmixed digesters were the earliest form of AD, operated at low loading rates are best suited for small facilities treating less than 3.8 m^3/h of sludge. At a typical detention time of 30 to 60 days, the content of these digesters is usually stratified with a scum layer on the top of the liquid surface and stabilised biosolid and grit on the digester floor. Mixed digestion takes place in AD with controlled mixing, heating; uniform feed rates to provide optimum condition for the microorganisms. The mixing systems can be gas mixing or mechanical mixing (Pump and Impeller).

AD can be operated in different stages depending on the nature of feedstock or the expected outcome form the digestion process. It could either be a single stage or two stage digesters. Continuous two-phase systems appear as more highly efficient technologies for anaerobic digestion of waste than the single stage (High rate) digesters (Bouallaguia, et al., 2005).

Table 2-2: Effect sludge concentration and HRT on OLR based on 70% volatile solid content and a sludge specific gravity of 1.02 (Eddy/AECOM, 2014)

Sludge concentration %	Organic loading rate (kg/m^3.d)			
	10d HRT	12 d HRT	15 d HRT	20 d HRT
2	1.4	1.2	0.95	0.70
3	2.1	`.8	1.4	1.1
4	2.9	2.4	1.9	1.4
5	3.6	3.0	2.4	1.8

6	4.3	3.6	2.9	2.1
7	5.0	4.2	3.3	2.5
8	5.7	4.8	3.8	2.9

However, due to the ease of operation of the high rate AD, it is most adopted as the process is controlled mainly by the sludge concentration, organic loading rate (OLR) and hydraulic retention time as shown in table 2-2.

2.4 Biochemical processes in AD

Every living organism undergoes some form of metabolism that consists of many biochemical pathways through which substrate molecules are converted into specific products needed by the organism. Biochemical pathways interact with one another so that cells have maximum flexibility in converting a wide array of substrates into specific molecules needed to function. There are two fundamental types of biochemical pathways: anabolic and catabolic. Anabolic pathways build complex products from less complex substrates while catabolic pathways break down complex molecules to make energy available for biological processes.

The biochemical pathway in anaerobic digestion process is thus catabolic and is carried out by various microorganisms consisting mainly archaea and bacteria. Figure 2-4 highlights the major biochemical processes that take place in AD.

In the process of AD, intracellular/extracellular enzymes act as facilitators while, solid retention time (SRT), hydraulic retention time (HRT), alkalinity, absence of toxic substances, bioavailability of nutrients and trace elements pH, temperature and gas concentration play very significant role in ensuring the process efficiency (Batstone, et al., 2002).

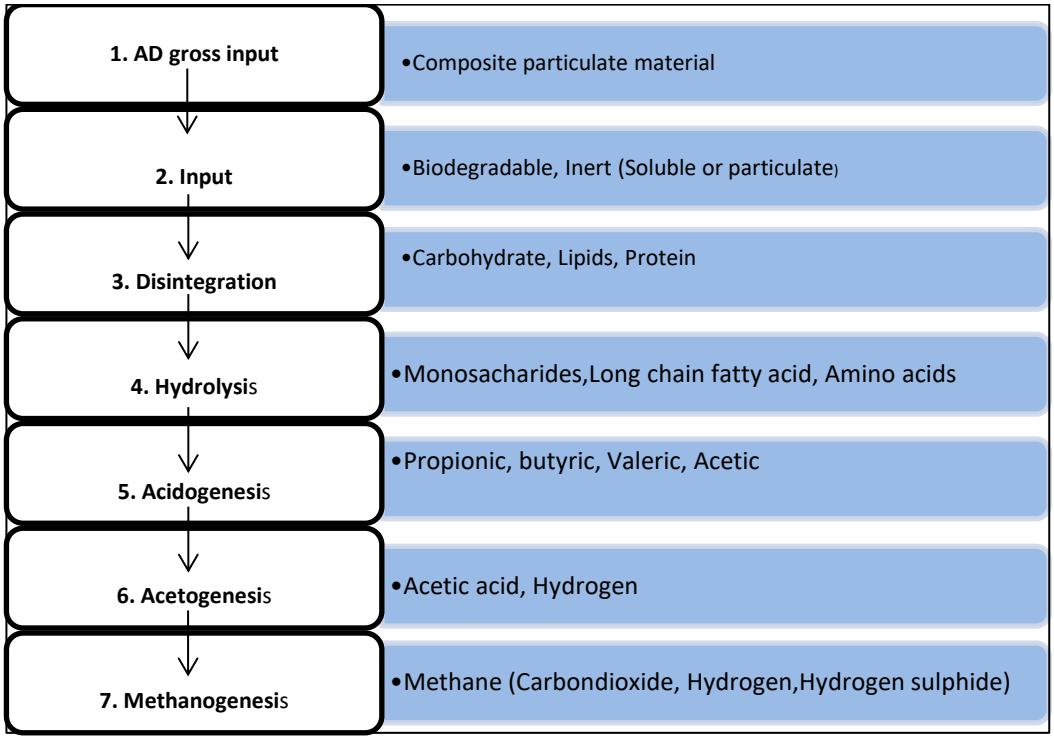


Figure 2-4: Summary of anaerobic digestion biochemical process

Referring to figure 2-4, the composite particulate material are usually PS, WAS, or a combination of both, with or without co-digestion which is the addition of other organic waste other than sewage sludge to the anaerobic digestion process

Co-digestion has been encouraged in recent times due to increased interest in biogas as a source of renewable energy. This is because of a higher percentage of solid content, volatile solid reduction (VSR) and consequently a higher volume of biogas in other organic waste compared with the separate anaerobic digestion of sewage sludge (Eddy/AECOM, 2014).

2.4.1 Disintegration

Higher VSR could be achieved in anaerobic digestion process through disintegration which is a non-biological process of breaking down lumps of composite particulate material. This can be realized through energy application in the form of ultrasound, heat, pressure or in combination. The methods include; ultrasound treatment, thermal hydrolysis, enzymatic hydrolysis pasteurisation and homogenisation (WEFTEC, 2007; Ometto, et

al., 2014). Ometto et al. (2014) concluded that enzymatic hydrolysis enabled the solubilisation of the cell wall constituents of microalgae pre-treated for improved AD compared to thermal and ultra sound pre-treatments. Amongst these disintegration methods for pre-treating organic waste and sludge, ultrasounds are most energetically imbalanced while thermal and enzymatic hydrolysis are the most energetically efficient (Ometto, et al., 2014).

2.4.2 Hydrolysis

Hydrolysis occurs when microbial population exudates hydrolytic enzymes such as lipase, protease and cellulose to break down soluble substrates (Jain et al., 1992, Vavilin et al., 1996). During hydrolysis, relatively pure substrates are broken down by the help of extracellular enzymes to particulate amino acids, monosaccharides, and long chain fatty acids (LCFA). Bacteroidetes and Clostridium have been found as the most prevalent microorganism responsible for hydrolysis in mesophilic and thermophilic AD systems (Strauber, et al., 2012) . The rate of hydrolysis can depend on many variables including pH, temperature, and the type of feedstock (Eddy/AECOM, 2014). These conditions affecting the hydrolysis stage make it the rate limiting step in the anaerobic digestion process as it is highly dependent on the ability of the microbial cells to naturally breakdown complex substance (Nicoletta, et al., 2014).

2.4.3 Acidogenesis

The acidogenic bacteria break down monosaccharide and amino acids to produce mixed organic acids such as propionic, butyric, valeric, acetic, ethanoic, etc., hydrogen and carbon dioxide. This process is known as acidogenesis and could involve different biochemical pathways depending on various factors such as pH, substrate concentration, dissolved hydrogen concentration and hydrogen partial pressure (Batstone, et al., 2002). For instance, at low hydrogen partial pressures the biochemical pathway that produces acetate and hydrogen is favoured compared to ethanol or butyrate formation thereby contributing to the main carbon flow from carbohydrates to methane formation. Acidogens consist mainly of the genera Bacillus, Bifidobacterium, Clostridium, Gracilibacter, Pseudomonas, Thermotoga and

Thermomonas as identified in mesophilic and thermophilic reactors treating wastewater (Balk, et al., 2002). Acidogens grow rapidly with minimum doubling times of about 30 minutes (Mosey, 1983).

2.4.4 Acetogenesis

The products of acidogenesis are further acted upon by acetogenic bacteria to convert LCFA, propionic, valeric, butyric, acetic, hydrogen and carbondioxide to acetate. Acetogenic bacteria comprise mainly of Genera such as Syntrophomonas and Syntrophobacter (Schink, 1997). The acetogenic bacteria degrade organic acids through oxidation devoid of any internal electron acceptor but require hydrogen ion or carbon dioxide to produce acetate (Batstone et al., 2002a) thus they maintain a syntrophic relationship with hydrogenotrophic methanogens. In addition, acetogenic bacteria are obligate hydrogen producers that are more efficient and thermodynamically favourable ($\Delta G' < 0$) at low hydrogen partial pressure.

$$2CH_3CH_2COOH + 2H_2O \rightarrow 4CH_3COO^- + 4H+ + 4H2\ (\Delta G°\ 96\ kJ/mol)$$

The acidogenic and acetogenic bacteria exhibit higher growth rates compared to the methanogens due to their ability to sustain on a wider source of feed and environment. The resultant effect is that at high organic load of readily degradable feed to the AD, the acidogenes and aceteogens proliferate rapidly as they consume the hydrolysed substance generating a higher concentration of acid in the AD. This sudden change in the environmental condition of the digester as it moves from neutral to acidic have a great impact on the activities of methanogens limiting the conversion of acetate to methane which then accumulate in the digester and inhibits the digestion process.

2.4.5 Methanogenesis

With additional activities of hydrogen utilising methanogenic and aceticlastic methanogenic groups of bacteria, methane is formed (Batstone, et al., 2002). The aceticlastic methanogens (AM), split acetate into methane and carbon dioxide while the hydrogenotrophic methanogens (HM) use hydrogen as the electron donor and CO2 as the electron acceptor to produce methane. As shown in the equation below (Batstone, et al., 2002);

$AM: CH_3COOH \rightarrow CH_4 + CO_2\ (\Delta G\ -31.0\ KJ/mol)$

$HM:\ 4H_2 + CO_2 \rightarrow CH_4 + 2H_2O\ (\Delta G\ -135.0\ KJ/mol)$

Homoacetogenic bacteria catalyse the inter-conversion between hydrogen and acetate which is a vital activity in the methane formation pathway. This could be achieved either by oxidizing or synthesizing acetate depending on the external hydrogen concentration. Aceticlastic methanogenesis function independent of hydrogen partial pressure while hydrogenotrophic methanogens function optimally at high partial pressure. The degradation of propionate and butyrate to acetate is not energetically feasible because the reaction is endergonic under standard conditions. Moreover, Ahring et al. (1995) identified that the accumulation of butyrate and propionate occur mostly during digester upset. Thus, the reaction is highly dependent on the presence of HM as shown in the equation below (Björnsson, 2000; Schink, 1997) reflecting the syntrophic relationship between acidogen and HM.

$CH_3CH_2CH_2COO^- + 2H_2O \rightarrow CH_3COO^- + H^+ + 2H_2$ (+48.1kJ/mol)

$CH_3CH_2COO^- + 3H_2O \rightarrow CH_3COO^- + H^+ + HCO_3^- + 3H_2$ (+76.1kJ/mol)

HM was found to be less sensitive to environmental changes than acetoclastic methanogens (Björnsson, 2000). However, Boe (2006) noted that at higher temperatures (> 30°C) the acetate oxidation pathway by homoacetogenic bacteria becomes more favourable. This compel the design and operation of AD to best suit the requirements of the methanogens. Hence, most AD is designed as mesophilic (30 – 38^0C).

Consequently, the methanogenic bacteria make up a very low proportion of the microbial community present in the AD. Although, there are seven known orders of methanogens: *Methanobacteriales, Methanococcales, Methanomicrobiales, Methanopyrales, Methanosarcinales, Methanocellales Methanomassiliicoccales* (Borell, et al., 2014), most of methane production arises from acetic acid conversions by acetoclastic methanogens such as *Methanosaetaceae* and *Methanosarcinaceae* mostly found in mesophilic and thermophilic AD. Only a small proportion of the methane production come from hydrogen/carbon dioxide reduction driven by hydrogenotrophic

methanogens including *Methanobacteriales, Methanomicrobiales*, *Methanococcales* and some members of the *Methanosarcinaceae* that can utilize both acetate and hydrogen/carbon dioxide to produce energy (Boone, et al., 1993).

2.5 Factors affecting anaerobic digestion process

Due to the complexity of the anaerobic digestion process, the performance of an anaerobic digestion process is broadly affected by many factors emanating from either sludge feed characteristics, digester process related, digester operating conditions or digester configuration. These broad categories are further discussed as follows;

2.5.1 Temperature

As in all biological processes, microorganisms involved in anaerobic processes perform optimally at specific temperature range. Temperature does not only affect the kinetic behaviour of the microorganism but also the thermodynamic and physical-chemical features such as viscosity and surface tension.

Most important is the effect of temperature on the growth rate and activity of the methanogenic organisms. Within ideal temperature range for a species of microorganism, the growth rate increases exponentially with temperature following the Arrhenius equation. A further increase beyond the optimal temperature does not result in digester improvement as this would alter the macromolecules and subsequently impede their metabolism.

On the other hand, lowering the temperature slows the reaction process. At approximately 10°C, digestion almost ceases as the methanogens are most effective in the mesophilic range (30 to 38 °C). Notwithstanding the specified digester operating temperature, the tank temperature must not deviate from that specified temperature value by more than 0.6 °C per day (Eddy/AECOM, 2014). This is essential to maintaining a large stable population of methanogen as each group has an optimum temperature for growth as shown in figure 2-5. Thus, the growth rate undergoes an exponential decline if the optimum temperature is exceeded.

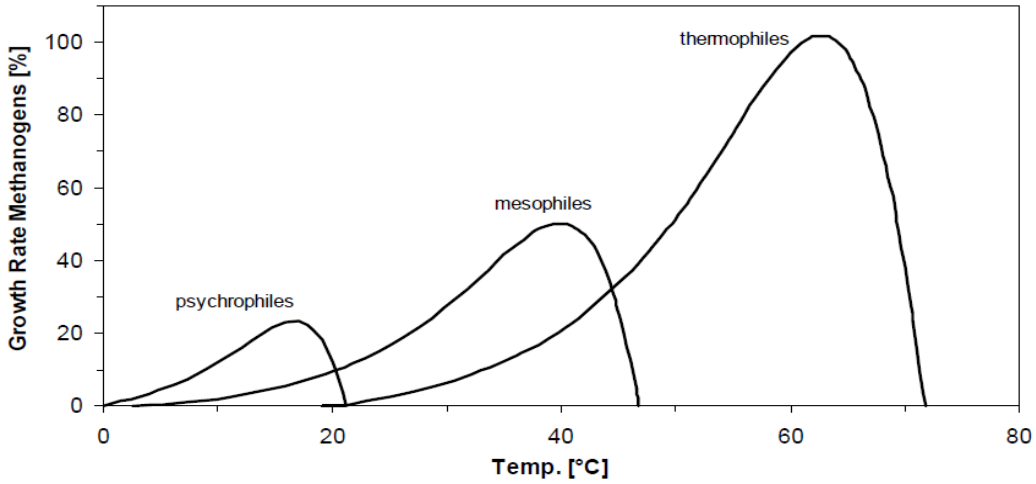


Figure 2-5: Relative growth rate of psychrophilic, mesophilic, and thermophilic methanogens, adapted from Van Lier etal; (1997)

In a psychrophilic anaerobic digester, methanogenesis occurs at a low rate requiring longer solid retention time (SRT) and organic loading rate OLR). On the contrary, thermophilic AD features a higher temperature leading to higher OLR, shorter SRT, faster microbial growth, and higher methane yield and pathogen destruction. However, most thermophilic AD face the challenge of maintaining the biomass resulting to incomplete degradation and low process stability.

Batstone et al.; (2002) identified that changes in physico-chemical parameters such as viscosity, equilibrium coefficients etc. and their overall effects on the AD are generally more important than those due to changes in biochemical parameters. For example, increasing temperature decreases the negative logarithm of the acid ionisation constant (pKA) of ammonia at pH > 8, thereby, increasing the fraction of free-ammonia which is inhibitory to microorganisms. In addition, increasing temperature increases pKA of VFA, which increase its undissociated fraction, especially at low pH < 5 rendering thermophilic AD process more prone to perturbation (Björnsson, 2000; Boe, 2006).

Typical causes of temperature stress to anaerobic digester operation include: overloading solids, exceeding the instantaneous capacity of the heating system, and operating the digester outside its optimum temperature range.

Although most heating systems will eventually heat the digester contents to the operating temperature, the variation in temperature for some period of time must have had an impact on the digester efficiency.

2.5.2 pH, Alkalinity, and Volatile fatty acid

Most microbial activities perform optimally at specific range of substrate pH level. For example, anaerobic digestion can operate at pH range from 6.0 to 8.0; however, the desired operating range for AD is 6.8 to 7.2 as the growth rate of methanogens fall sharply at pH less than 6.8. The acidogens functions best within pH ranges from 8.5 to 4 (Hwang et al., 2004) with an optimum around 6 while the acetogens operate optimally at a pH around 7 (Eddy/AECOM, 2014).

Alkalinity is a measure of the buffering capacity of the digester. It is principally the salts of weak acids and strong bases usually expressed as concentration of calcium carbonate with an optimal range of 3000 to 5000 mg/L as CaCO3 (Eddy/AECOM, 2014). The presence and concentration of a buffering compound depends on the composition of the substrate and the total organic load. For example, through the destruction of digester feed containing nitrogen, bicarbonate alkalinity is produced as the released ammonia nitrogen reacts with carbon dioxide (Grady et al., 1999) as shown in the chemical equation (Eddy/AECOM, 2014)

$NH_3 + H_2O + CO_2 \rightarrow NH_4+ + HCO^{-3}$

Possible substance that could be used to boost the alkalinity and their equivalent weight ratios are as shown in table 2-3

Table 2-3: Potential alkalinity and their equivalent weight ratios adapted from (Eddy/AECOM, 2014)

Chemical	**Ratio**
Anhydrous ammonia (NH_3)	0.32

Aqua ammonia (NH_4OH)	0.7
Anhydrous soda ash (Na_2CO_3)	1.06
Caustic soda (NaOH)	0.8
Hydrated lime ($Ca(OH)_2$)	0.74

Lime addition can be messy and will produce $CaCO_3$ while ammonia toxicity and increase in ammonia load on the liquid treatment processes through return streams may arise if ammonia is used. Applying alkalinity over a 3- or 4-day period, mixing well, and monitoring volatile acids, pH, and alkalinity will help to avoid toxicity from the cations associated with the alkalis (Eddy/AECOM, 2014).

On the other hand, volatile acid is the measure of acidity of the anaerobic digestion process. Mostly, it is measured as the concentration of acetate with an optimal range of 50-300 mg/L for methanogenic AD treating sewage sludge. Volatile acid present in AD is made up mostly of volatile fatty acid (VFA) and carbonic acids (Eddy/AECOM, 2014). The VFA are intermediate products (metabolites) of the anaerobic digestion of organic matter which include mainly acetic, propionic, butyric, valeric and sometime ethanoic acid (Ferguson, et al., 2016). Biochemical pathways yielding acetic acid are of great importance in AD as greater than 70% of methane generated in AD is from acetic acid (Wijekoon, et al., 2011).

Although VFA have been identified as vital indicator of digester upset, there have been contradictions on the combination of VFA that best represent the onset of digester upset. Angelidaki et al. (1993) identified the difficulty in establishing VFA levels that indicate the state of AD thus they concluded that VFA for different system varies based on their buffering capacity. To further illustrate this point, Wang et al. (2009) reported that acetic acid and butyric acid concentrations of 2400 and 1800 mg/L respectively, resulted in no major

methanogenic inhibition, while propionic acid concentration of 900 mg/L resulted in major methanogenic inhibition.

Consequently, as evident from figure 2-6, initial accumulation of VFA are neutralised by the bicarbonate alkalinity present in the substrate as illustrated for acetic acid (HAc) in the following equation:

$$HCO^{-3} + HAc \rightleftharpoons H_2O + CO_2 + Ac^-$$

However, once there is an imbalance in the system with a slight increase in the VFA concentration the methanogens activities are inhibited, and carbon dioxide and hydrogen accumulate in the digester. In addition, VFAs will react with bicarbonate alkalinity, both reducing its concentration and producing carbon dioxide, which increases the carbon dioxide content of the gas space. Consequently, carbon dioxide solubilises and reacts reversibly due to the partial pressure of gas in a digester as shown in the equation to form carbonic acid. (Bischofsberger et al., 2005; (Eddy/AECOM, 2014)):

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

This highlights the fact that the principal consumer of alkalinity in a digester is carbon dioxide, and not solely VFA (Eddy/AECOM, 2014). However, higher concentrations of volatile acids are acceptable in AD if sufficient alkalinity is available to buffer the acid. As shown in figure 2-6, appropriate level of alkalinity is required to buffer the drop in pH due to increase in a combination of both carbonic acid and VFA. Manure and sewage fed digesters normally have high alkalinity and ammonia content thus enabling the digester to maintain a stable pH of 7.5-8.0 (Boe, 2006).

This was evidenced by Y.Bajon Fernadez et al. (2014), when investigating carbon dioxide enrichment of AD in food waste and sewage sludge. Methane production was enhanced by up to 13% and 138% for digesters treating food waste and sewage sludge respectively. The results demonstrated the ability of AD to use additional carbon dioxide as long as there is sufficient alkalinity to buffer the system and enhance acetoclastic pathway of methane formation.

Figure 2-6a

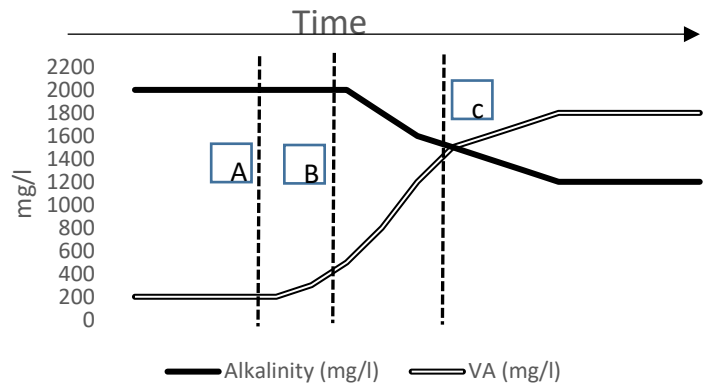


Figure 2-6a shows a digester operating with a good buffering capacity (Volatile acid (VA)= 200 mg/l). At point A, an event in the AD resulted to an increase in VA followed by a decrease in alkalinity at point B. At point C the digest has

Figure 2-6b

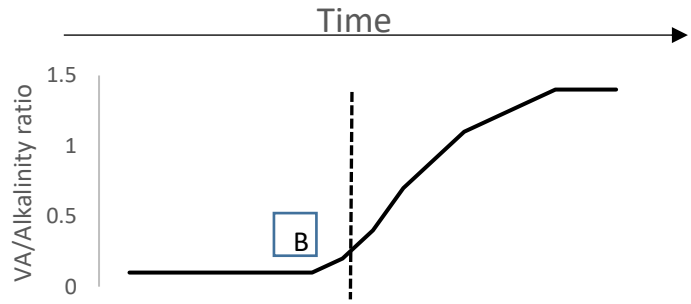
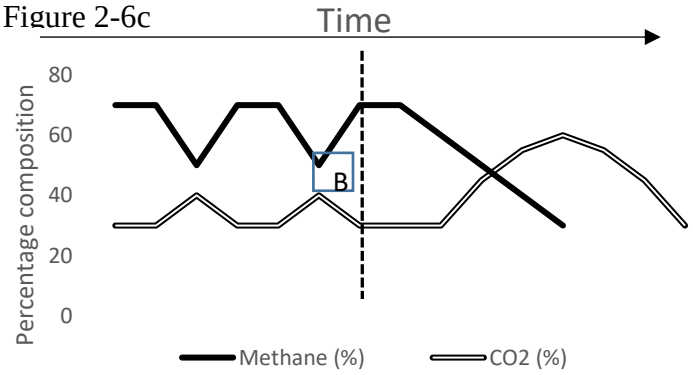


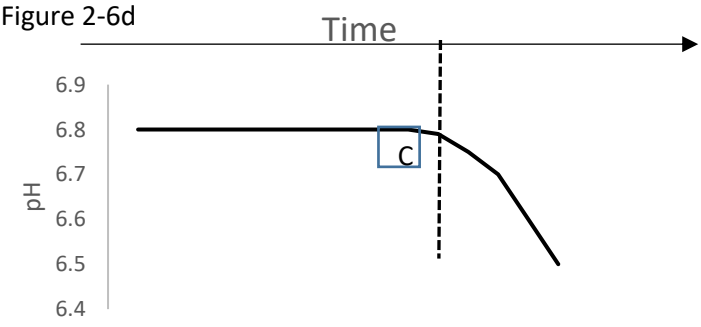
Figure 2-6b shows the same digester referencing the VA/Alkalinity ratio. At point B, the increase in VA produces an increase in the ratio from 0.1 to 0.3.

Figure 2-6c



By comparing figure 2-6c with Figure 2-6b; methane production begins to drop with a corresponding increase in carbon dioxide when the ratio in figure 2-6b reaches about 0.5

Figure 2-6d



pH does not change in figure 2-6d until the digester is becoming sour at point C

Figure 2-6: Sequential changes in digester due to variation in operating condition (adapted from WEFTEC, 2007)

Thus, VA/Alkalinity >0.8 depresses the digester pH and inhibit methane production. However, VA/Alkalinity > 0.3 to 0.4 point out the tendency of digester upset requiring corrective action even before a pH change occurs. Usually, corrective action should be taken by adding alkalinity and decreasing the volatile solids loading rate of the AD by ensuring an approximately 25% decrease in VS feed and sludge withdrawal rates as well as maintaining the internal digester sludge temperature at 35 °C for mesophilic digesters (Eddy/AECOM, 2014).

Varying methods for analysis of VFA have evolved over the years due to the significance of VFA in controlling the anaerobic digestion process. These method for VFA analysis include inter alia; colorimetric (Montgomery et al. 1962), titrimetric (Mota, et al., 2015), high pressure liquid chromatography (HPLC) (Sansone & Martens, 1981; Vasconcelos de Sáa, et al., 2011) ion chromatography (IC) (Manning & Bewsher, 1997), and gas chromatography (GC) (Siedlecka, et al., 2008).

In recent time, the trend in industrial revolution through advancement in computing system and adoption of integrated sensor system has led to the development of on-line sensors for VFA monitoring such as the OPTI-VFA sensor (Deng, 2015). OPTI-VFA was developed based on Fabry-Perot spectrometer, ATR probe and automatic control of AD process to provide real-time measurements of VFA (Deng, 2015). GC analysis of VFA was the reference measurement applied in developing and validating the calibration models of the OPTIVFA measurement systems. Chemometric models were developed to predict VFA concentrations based on FT-IR/ATR spectral data.

Validating the performance of the OPTI-VFA measurement and control of AD process was carried out using 2 stage mesophilic AD treating bio-waste (garden waste and kitchen waste from households) in laboratory and full scale (Deng, 2015). Although there are existing reference analytical methods, OPTI-VFA is unique as it requires no sample preparation. The control identifies the status of the AD reactor using the probe and automatically adjusts the AD process to achieve maximum treatment capacity. Applying a simulation based procedure, OPTI-VFA was designed to be able to

simultaneously deal with the dimension of the plant components such as reactor volume and gas holder.

Considering that the application of OPTI-VFA control was best designed for a 2-stage AD whereby higher percentage of feed hydrolysis and acidification is achieved in the first digester and methanogenesis progresses in the second stage of the digestion process. It is evident that the application of such technology for AD treating sewage sludge which is significantly operated as a single stage AD will need some modification of the control system to be more suitable in efficiently controlling AD treating sewage sludge.

2.5.3 Inoculum/Substrate composition (nutrients and trace elements)

The objective of adequately planning and executing anaerobic digester start-up is to achieve steady-state operation and the required volatile solids reduction in the shortest possible time. During the preparation phase, typically the first 4 days, the digester is filled to its minimum operating level, allowed to reach operating temperature before mixing and feeding is started (Eddy/AECOM, 2014). This initial filling of the digester can be achieved using any of the following liquids that serve as inoculum: unthicken WAS prior to polymer addition, unchlorinated secondary clarifier effluent, primary clarifier effluent, and combination of PS and WAS.

Inoculum is not essential for effective AD start up however, it could reduce the AD start-up from the normal 45 days to 7 days if adequate and active inoculum is used (WEFTEC, 2007). The ratio of primary solids to WAS in the inoculum should resemble the digester feed as closely as possible.

The nature of the substrate fed to the digester affects the overall efficiency in terms of steady operation, biogas quality and volatile solid reduction. Greater AD efficacy could only be achieved if the microorganisms involved in AD operation could adequately source energy (carbon) to carryout synthesis of new cellular material and growth factors (inorganic elements /nutrients) to proliferate and function properly (Eddy/AECOM, 2014). The composition and biodegradability of the digester feed affects the chemical conditions in the reactor. Readily biodegradable substance such as

carbohydrate will result in an increment in the available VFA for methanogens as the hydrolysis step will occur faster than if it were protein. This will in turn lead to a change in digester pH and may result in process inhibition.

Report from WRAP (2010) on food waste composition identified up to 50% variation in the carbohydrate content of food waste in the winter compared to the summer. In wastewater treatment, variation in the quality and quality of activated also depend on the season as well as dependent on the upstream process. Thus, if there is a failure in the different stage of activated sludge treatment or thickening process, then it will impact greatly on the digester performance.

Feedstocks to AD are often measured in terms of total chemical oxygen demand (COD) or total volatile solids (VS). It is very crucial to identify the amount of solid present in the feedstock that is biodegradable (VS) in order to effectively manage the impact on the system and have a fair estimate of the grit accumulation in the digester and when it will be necessary to dispose of them so that it does not unnecessarily take up the digester space.

Furthermore, some nutrients found in the feedstock could be very useful for adequate digester operation at certain concentration. However, above these concentrations, they could become inhibitory. Efficient biodegradation requires that carbon sources and nutrients are available in sufficient amounts in the substrate. Common ration used in accessing suitability of the digester to function optimally include; C:N = 10:1 to 30:1, N:P = 5:1 to 7:1 as well as COD:N:P =420:7:1 to 1500:7:1 (Bischofsberger et al., 2005; Eder and Schulz, 2006; Gray, 2004 as referenced in (Schön, 2009)).

The main chemical composition of AD feed includes carbon, hydrogen, oxygen and nitrogen and in some instances sulphur. Using the formula by McCarty, the conversion of organic matter to methane could be determined;

$$C_nH_aO_bN_c + \left(2n + c - b - \frac{9sd}{20} - \frac{ed}{4}\right)H_2O \rightarrow \frac{de}{8}\,CH_4 + \left(n - c\,\frac{sd}{5} - \frac{de}{8}\right)CO_2 + \frac{sd}{20}\,C_5H_7O_2N + \left(c - \frac{sd}{20}\right)NH_4^- + \left(c - \frac{sd}{20}\right)HCO_3^-$$

Where;

d = 4n + a -2b-3c

s = fraction of waste converted to cells

e = fraction of waste converted to methane for energy (s + e =1)

$C_nH_aO_bN_c$ = empirical formula of waste being digested

$C_5H_7O_2N$ = empirical formula of bacterial dry mass (VSS)

McCarthy also proposed a rule of thumb for estimating methane yield from waste stabilisation; 0.45 kg of COD stabilised equals $0.16m^3$

Analysis of previous studies on inoculum to substrate ratio by Eskicioglu & Ghorbani (2011) identified that the ultimate methane yield and methane production rates are feedstock and substrate specific thus could hardly be generalised.

2.5.4 Toxicity

Inhibition occurs through prevention or reduction of the enzymatic activities on the substrate which could be competitive or non-competitive (Eder and Schulz, 2006). Competitive inhibition of enzymatic activities occurs when the attachment of the inhibitor to the enzyme prevents the substrate from attaching to the enzyme. This can easily be managed if the concentration of the substrates is further increased such that it reduces the chances of the inhibitor accessing the enzyme and lessening its impact on the digestion process. However, when the inhibitors attachment to the enzyme is capable of destroying the cellular content of the enzyme, this is a non-competitive inhibition which results to an irreversible process hindering the digester restoration to normal operating condition (Boe, 2006).

Sources of toxic substance could be from the inoculum, feed or by products from the breakdown of the digester feed. Most common inhibitors are formed during substrate degradation and include VFA, LCFA, and ammonia. Under normal circumstances, these substances do not inhibit the digester operation,

however, variation in physio-chemical condition in the digester could make them inhibitory such as; pH, organic loading, temperature, hydraulic loading, the presence of other materials, hydrogen partial pressure and the ratio of the toxic substance concentration to the biomass concentration.

Toxicity to microbial organism due to ammonia and VFAs is highly dependent on pH as only the non-ionised forms contribute to toxicity. At the optimal pH range for anaerobic digestion (6.8 – 7.2), ammonia is mainly present as NH_4^+ thus does not constitute any inhibition to the digestion process. In the same manner, VFAs become inhibitory when the pH of the system has gone below 5.0 (Batstone et al., 2002) by penetrating and crossing the cell membrane of the microorganism where a higher pH is maintained. Consequently, the non-ionised VFA is ionised, thereby releasing hydrogen ion which will cause a decrease in the intercellular pH as well as disrupting homeostasis (Eddy/AECOM, 2014). In addition, temperature affects the pH level and the surface tension thereby impacting on the proportion of non-ionised VFA and ammonia present in the digester. (Kanu, et al., 2015a)

At even low concentration, Longer Chain Fatty Acids (LCFA) and especially their isoforms is the intermediate product of lipids in the anaerobic digestion process and have been proposed to be inhibitory to most microorganisms (Fischer, et al., 1984). They adsorb onto the cell wall membrane thereby inhibiting the transport of essential nutrients (Batstone et al., 2002).

On the other hand, the anaerobic process is also sensitive to certain compounds, such as sulfides, volatile acids, heavy metals, calcium, sodium, potassium, dissolved oxygen, and chlorinated organic compounds. Trace elements such as heavy metals are essential nutrient for digester operation which are required in very low concentrations. They include the following; fluorine, iodine, chromium, manganese, zinc, nickel, cobalt and copper. When trace elements are not sufficient at the right concentration, this may impact negatively on the growth and activities of the microorganism whereas higher concentrations of such heavy metals are toxic (Bischofsberger et al., 2005; (Eddy/AECOM, 2014)). Sodium sulphide or ferric or ferrous sulphate can be added to alleviate heavy metal toxicity because toxic heavy metal sulphides

have low solubility and are less soluble than ferric sulphide, the toxic metals precipitate as sulphides.

Sulphide (S^{2-}) is a product of the degradation of influent sulphate ($SO4^{2-}$ which is not inhibitory) and sulphur-containing biodegradable substance such as protein. This is possible as sulphate serve as electron acceptors to sulphate reducing bacteria bringing about the reduction of sulphate to sulphide (Schink, 1997). Further conversion of sulphide to hydrogen sulphide at pH < 7 result to an increase in the corrosion ability of the gas and formation of sulphur oxides when the gas is combusted which destroys the components of the combustion engine. On the other hand, since hydrogen sulphide are not readily soluble in water, thus, they partition between the liquid and gas phases thereby inhibiting digester operation at concentrations > 20 mg/L and >5 mg/L in the gas and liquid phase, respectively (Eder and Schulz, 2006; Grady et al., 1999). Soluble sulphides become toxic at concentrations >200 mg/L. Ferric chloride can be used to control sulphide concentration by precipitation as ferric sulphide. Nevertheless, excessive use of these chemicals can result in pH reduction (Eddy/AECOM, 2014).

In addition, sulphate-reducing bacteria compete with methanogenic archaea for hydrogen and acetate at low concentration of sulphate while competing with acetogenic bacteria for propionate and butyrate at high concentration of sulphate. Due to the diverse metabolic capability of sulphate reducing bacteria, they exhibit competitive advantage over hydrogenotrophic methanogens for hydrogen (Stams et al., 2005). Boe (2006) reported the complete eradication of hydrogenotrophic methanogens because of competing with hydrogen-utilising sulphate reducers in an anaerobic reactor treating sulphate-rich wastewater.

Antibiotics present in wastewater sludge can also be inhibitory at high concentration to anaerobic digestion process. A lot of researches are currently on-going on quantification of impact of antibiotics on microbial activities in AD and possible measures of recovery.

Since, AD is designed to function in the absence of oxygen, the strictly anaerobic methanogens are very sensitive to the presence of oxygen in the

digester. However, within the mixed microbial population found in AD, the facultative bacteria will rapidly consume the oxygen thereby restoring the digester to its anaerobic status. (Björnsson, 2000; Eder and Schulz, 2006).

2.5.5 Internal mixing and retention time

Combining efficient mixing of AD systems with other operating condition such as temperature control and uniform feed rates is very beneficial to providing optimum environmental conditions for the microorganisms that accomplish anaerobic digestion. Effective mixing enables the AD system to achieve process stability, prevent solid/grit accumulation and control scum/foam nuisance.

Mixing provides process stability by ensuring that there is sufficient contact between the feed and active digestate such that the microorganism could sufficiently degrade the sludge. Mixing enhances gas release from the sludge and assist in maintaining a uniform temperature and solids mixture throughout the tank thus preventing solid accumulation. Poor mixing have been identified as the cause of stratification within the digester leading to withdrawal of partially digested sludge from the AD (Eder and Schulz, 2006; Gray, 2004)

Since scum accumulation is majorly on the liquid surface especially in locations of low turbulence, it is efficiently controlled by mixing the scum into the digester contents (Eddy/AECOM, 2014). Nevertheless, large-diameter tanks allow it to spread over a larger area, making it more difficult to combine. On the other hand, accumulations of grit deposit at the bottom of AD lead to loss of active digester volume. Cleaning up such grit deposit could be expensive and difficult with a lot of health and safety implication thus is best to avoid it. This could be achieved by adequate mixing and use of sloped bottoms such that grit and settleable solids slide down the sides to a single location where they can be either removed or lighter material re-suspended.

Issues relating to low mixing intensity such as poor contact of the biomass with the substrate, low activity of biomass going, and lower biogas production

are prevalent in psychrophilic which limit the ability of the AD to undergo natural mixing.

2.5.6 Digester and feed solid

The ability of the digester to efficiently degrade solids is highly dependent on the solid retention time (SRT) and organic loading rate (OLR). These variables control the amount of food the microorganisms must stabilize and the time available to consume that food. When the feed solids concentration is below 3%, SRT controls the processing capacity while OLR control the processing capacity when it is higher than 3% (Eddy/AECOM, 2014).

2.5.6.1 Solid Retention Time (SRT)

The solids retention time (SRT) equals the period that solids are retained in the digester and is a crucial factor affecting reactor performance. Typical well-mixed high-rate digesters encounter SRT in the range of 15 to 20 days. Eder and Schuz (2006) identified that the minimum SRT can be estimated based on generation times of the microorganisms involved in AD process and the type of digester feed as shown in figure 2-7.

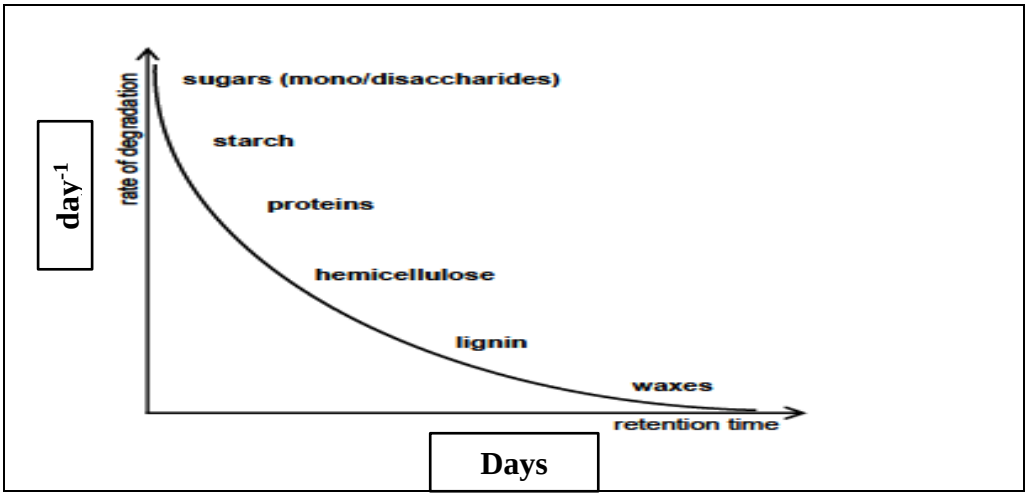


Figure 2-7: Rate of degradation of different digester feed against retention time (adapted from Eder and Schuz, 2006)

There will be washout of most microbial population and digester solid if SRT is shorter than the generation time of the slowest group of microorganisms which are normally the acetogens. Thus, SRT controls the different types of

microorganisms growing in AD thereby influencing the degree of degradation as well as the gas yield (Grady et al., 1999).

Consequently, short retention times provide high gases production rates (related to the reactor volume) with low volatile solid destruction as the easily degradable substrates are digested wasting the non-easily degradable solids.

HRT equals SRT for completely mixed AD operating without solids recycling. When solid recycling is involved in the AD process, SRT will be greater than the HRT (Grady, et al., 1999). In addition, shorter retention times also reduce the system's buffering capacity (Eddy/AECOM, 2014).

2.5.6.2 Organic loading rate

To satisfy process requirements, the digester contents must not be subjected to large fluctuations in raw sludge substrate, solids concentration, and bacterial metabolic end products. To ensure that the digester volume is effectively utilised, while maintaining these conditions, OLR is used to determine the loading to the digester. It can be expressed in terms of the mass of volatile solids applied as shown in equation:

$$OLR = \frac{Q \times C}{V} = \frac{C}{HRT} \qquad \text{Equation 2-1}$$

where

OLR = the volumetric organic loading rate (kgVS/ m^3d),

Q = influent flow rate (m^3/d),

C = concentration of volatile solids in the digester feed (kgVS/ m^3)

V = bioreactor volume (m^3)

HRT = V/Q (d)

Although in some instance, chemical oxygen demand or total solid have been used to estimate digester loading, the volatile solid tends to account better

for the biodegradable content of the digester thus should be a more appropriate measure of digester organic load.

During digester start up, control of the OLR is very critical to efficient digester operation. The initial OLR should be approximately 0.16 kg volatile solids/m3 d and should be increased every 3 days, provided that the digester process-control parameters are within the desired limits as shown in table 2-4 (WEFTEC, 2007).

Table 2-4: Major conditions to be monitored for the efficient start up and operation of mesophilic anaerobic digester.

Parameter	Target	Rate	Sampling location
Temperature (°C)	32 - 38	Daily	Digestate
Volatile acids (VA) (mg/ L)	50 - 330	Daily	Digestate
Alkalinity (mg/ L)	1500 - 5000	Daily	Digestate
VA:Alkalinity	0.1 – 0.2	Daily	Not available
pH	6.8 - 7.2	Daily	Sludge feed
Total solids (%)	Monitor the trend	Daily	Sludge feed
Volatile solids (VS) (%)	Monitor the trend	Daily	Sludge feed
Organic loading rate ($kgVS/m^3d$)	1.6 – 3.2	Daily	Sludge feed
Gas production (m^3/ kg of VS destroyed)		Daily	Gas storage
Gas composition	< 35% CO_2	Daily	Gas storage

Regular feeding of the digester should be carried out to curb the instantaneous loading rates on the digester and minimize foaming potential. The operating parameters in table 2-3 should be monitored daily during start up and analysed using trend charts to identify both positive and negative trends and predict possible digester upsets. This is crucial as within the target range, the rate of change of an individual parameter is more significant than its absolute value (Kanu, et al., 2015a).

Hence, OLR should be increased every 3 days at increments of 0.16 $kg\ of\ volatile\ solids/m^3d$ once the operating parameters indicate that the digester is operating at steady state until the design volatile solids loading rate is attained. R.M.W Ferguson et al. (2016), operated AD at 38 ℃ with a

7-day retention time while investigating the effect of changes in OLR as a promising microbial management tool in AD.

The maximum possible OLR depends on the process retention time and the type of digester (temperature). Lower temperature and longer retention time result to higher OLRs. There have been various report on the ideal OLR for AD treating sewage sludge such as; 1.6 to 3.2 $kg\ of\ volatile\ solids/m^3d$ (Eddy/AECOM, 2014) and 2 to 6 $kg\ of\ volatile\ solids/m^3d$ (Eder and Schulz, 2006; Grady et al., 1999). These ranges vary for other waste stream and could be higher.

Organic over loads have an adverse effect on digester operation and could be attributed to any of the following; rapid digester start up, excessive digester feed exceeding the design volatile solid limit by 10% and loss of active digester volume due to solid/grit accumulation arising from poor mixing (Eddy/AECOM, 2014). Feeding the system above its designed OLR will result in low biogas yield due to accumulation of inhibiting substances such as fatty acids in the digestate.

A study by Ferguson et al (2016) identified that digesters exposed to one or two changes in OLR using the same or different co-substrates (Fat Oil and Grease waste (FOG) and/or glycerol produced a decrease in biogas and methane production. However, the digesters exposed twice to glycerol showed faster recovery towards stable conditions after the second OLR change. Since quality of feed to AD fluctuate throughout the year, it is often difficult for an operator to maintain stable digester OLR hence the need to develop efficient biotechnological tools to manage such complexity.

2.5.6.3 Volatile solid reduction (VSR)

The percentage reduction in volatile solid is often used to establish the extent to which the digester has stabilised in addition to biogas production rate/quality, alkalinity and volatile fatty acid concentration (Eddy/AECOM, 2014). Liptak (1974) proposed an empirical equation for the rough estimation of volatile solid destruction in a complete mixed digester as shown in

$$VSR = 13.7\ ln\ (SRT) + 18.9$$ Equation 2-2

Where;

VSR = Volatile solid reduction in %

SRT= Sludge retention time in days (12 to 20days)

Eddy/AECOM (2014) identified that the equation does not account for differences in sludge feed and other digester operating conditions, thus, leading to an overestimate of VSR.

Using the method of mass balance as shown in equation 2-3, it is assumed that the ash content that entered the digester is equal to the ash content leaving the digester (Eddy/AECOM, 2014). Thus, VSR equals the ratio of loss in volatile solids to sum of volatile solid inputs.

$$\text{VSR} = \frac{M_f - M_d - M_s}{M_f} \times 100$$ Equation 2-3

Where; M_f = Mass flowrate of volatile solid in the feed sludge in kg/d

M_d = Mass flow rates of volatile solid in the digestate in kg/d

M_s = Mass flow rates of volatile solid in the supernatant in kg/d

However, in modern high rate digesters, there is no removal of supernatant thus M_s equals zero and there is tendency that the proposed ash content accumulates in the digester in the form of grit. In addition, the mass of volatile solid is determined per unit mass of total solids.

Notwithstanding the perceived simplicity in calculating the VSR using the mass balance method, in practice, this could be quite challenging as it

involves the determination of all feed and withdrawal volumes, initial and final volumes in the digester and determination of volatile solids concentration on all streams (USEPA, 1989). Hence, there is need for the process to be simplified.

Consequently, Van Kleeck formular was established as in equation 2-4 (WPCF, 1968). Although it is not valid when there is grit accumulation as it assumes that the fixed solid input equals a fixed solid output, however, it produces a conservative result. Since it does not account for non-removal of supernatant, which is the most applicable scenario in most AD operation, this Van Kleeck formula is widely adopted by most operators of full scale AD.

$$\text{VSR} = \frac{W_f - W_d}{W_f - (W_d)(W_f)} \times 100 \qquad \text{Equation 2-4}$$

Where;

W_f = Weight fraction of volatile solid in feed sludge per total dry solid

W_d = Weight fraction of volatile solid in digested sludge per total dry solid

2.6 Foam

Most foam occurs as a medium of gas trapped in thin fluid film. Basically, foams are unstable and will likely collapse to a liquid which is its lowest energy state. It has been noted by Guitian & Joseph (1998) that the intense adsorption of surfactants at the walls of the bubble opposes the collapse. The surfactants adsorb on the bubble walls as they are both hydrophobic and hydrophilic. The hydrophobic sect which often has an organic structure tends to move away from the fluid while the hydrophilic sect which is either polar or have charges that separate when dissolved in water moves toward the fluid forming (Guitian & Joseph, 1998). This primarily lowers the surface tension of the liquid thus easing the formation of a more stable foam.

Foam has an effective yield stress as it can trap and restrain the movement of small and light particles, although, these entrapped particles cannot mix

up in the foam. Foam possesses some level of viscosity for flow but does not mix up like fluids. The fall can illustrate the viscoelastic property of foam out of heavy particles driven into the foam in a chain of interconnected particles (Guitian & Joseph, 1998). Foam can further be described in terms of the ease with which the foam is formed, and its stability determined by how long it took for the foam to collapse to a liquid. Therefore, foam can be categorised as either stable, metastable, or unstable (Westlund, et al., 1998).

2.6.1 Foam formation, stability, and collapse

Gas-liquid-particle mixtures (such as those found in AD) will foam only when surfactants are present to lower the surface tension of the liquid and trap gas bubbles as they blow up. Thus, chemically pure liquids do not foam. Usually, the developed foam is made up of about 80% gas. Everyday foaming experience ranges from soap bubbles to beer bubbles as it is poured into a glass. In each case there is presence of, liquid (surfactant) and bubbles (gas).

Foam stability and collapse are very crucial in foam life as they determine the intrinsic qualities of foam. It is difficult to explain foam stability separately from foam collapse as they are two inevitable conditions. This means that, foam when formed will either stabilise or collapse. Basically, foams are unstable and will likely collapse to a liquid which is its lowest energy state (Guitian & Joseph, 1998).This usually occurs because foam film though seemingly stable, will bursts at hundreds of centimetres per second, initiating a visible rearrangement of bubble assembly through surface and fluid forces occurring over less than a second (Saye & Sethian, 2013).

It has been noted by (Guitian & Joseph, 1998) that the intense adsorption of surfactants at the walls of the bubble opposes the collapse of foam. The surfactants adsorb on the bubble walls as they are both hydrophobic and hydrophilic. The hydrophobic sect which often has an organic structure tends to move away from the fluid while the hydrophilic group which is either polar or have charges that separate when dissolved in water moves toward the fluid forming (Guitian & Joseph, 1998). In addition, an increment in surface tension is observed whenever a film is suddenly stretched locally (such as

mixing), resulting in increased opposition to the destabilising force (Weaire, et al., 2003)

Coarsening is the gradual change of the foam structure due to gas diffusion through the films which is dependent on pressure differences between bubbles. Small bubbles have high pressure and disappear quickly as they lose gas to bigger bubbles. As this happens, average foam size increases with time as foam coarsens.

The ratio of liquid content to foam is very important when considering foam stability. Liquid fraction in foam can range from zero to about 35%, the wet limit, at which the bubbles come apart (Weaire, et al., 2003). Whenever there is a set amount of liquid in the foam, it tends to drain due to the influence of gravity and local variations in pressure through films, Plateau borders and junctions of the static foam structure. This is in accordance with the hydrostatic pressure law necessary for equilibrium under gravity and also depends on the type of surfactant dominating in the liquid (Guitian & Joseph, 1998).

Disjoining pressure is essentially the mutually repulsive force between two faces of a film which opposes further thinning. The film thickness at which equilibrium is achieved is determined by the balance between the disjoining pressure and the bulk pressure of the liquid. The disjoining pressure increases as distance between the films decreases while the bulk pressure decreases to large negative values as the liquid fraction of the foam decreases (Weaire, et al., 2003). Thus, if thinning is opposed, then the rate at which the foam collapse declines.

On the other hand, the presence of particles has been shown to accelerate foam collapse as shown in figure 2-8. Guitian & Joseph (1998) found that fluidized solid particles are at zero order, thus acting as stationary objects over which liquid could pass. In addition, solid particles increase the effective density of the gas bubble thereby compelling the gas to rise at a faster velocity causing a decrease in gas holdup as buoyancy is proportional to the difference between the gas density and the density of the liquid plus solid mixture (Guitian & Joseph, 1998). This provides less time for

emulsification of the bubbles by surfactants resulting in less foam formation and stability.

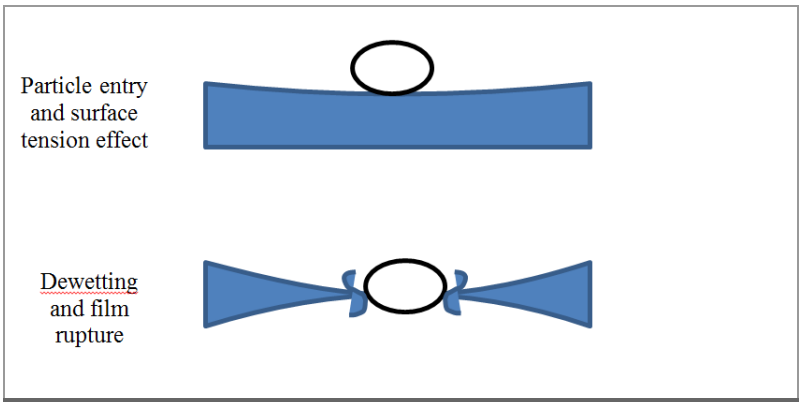


Figure 2-8: Effect of solid particles on foaming

2.6.2 Foaming in anaerobic digester

Ganidi, et al (2009) defined AD foaming as a built up of a dark mixture of gas bubbles bounded by liquid films on the surface of the sludge while in activated sludge wastewater treatment process, microbial foam appears as a dark brown viscous layer (Oerther, et al., 2001). Formation of foam is one of the major causes of process upset in biogas plants (WEFTEC, 2006). No one knows the exact cause of foaming as each sludge and digester design is a unique combination and presents a unique set of circumstances with foaming occurring under the 'best' of circumstances (Williams & Tim, 2012).

However, foaming is considered excessive if it blocks piping and/or escapes the containment of the anaerobic digester (AD) (WEFTEC, 2006). The tendency for a foaming episode to constitute nuisance is dependent on the stability of the foam. As can be seen from figure 2-9, the first column on the right is experiencing stable foam as there is a gradual and sustained foam level for a long period of time. On the contrary, metastable foaming is taking place in the second column by the right with foaming level not rising nor reducing so rapidly. The columns on the left experienced unstable foaming resulting in rapid rise and fall of foaming episode as shown in figure 2-9.

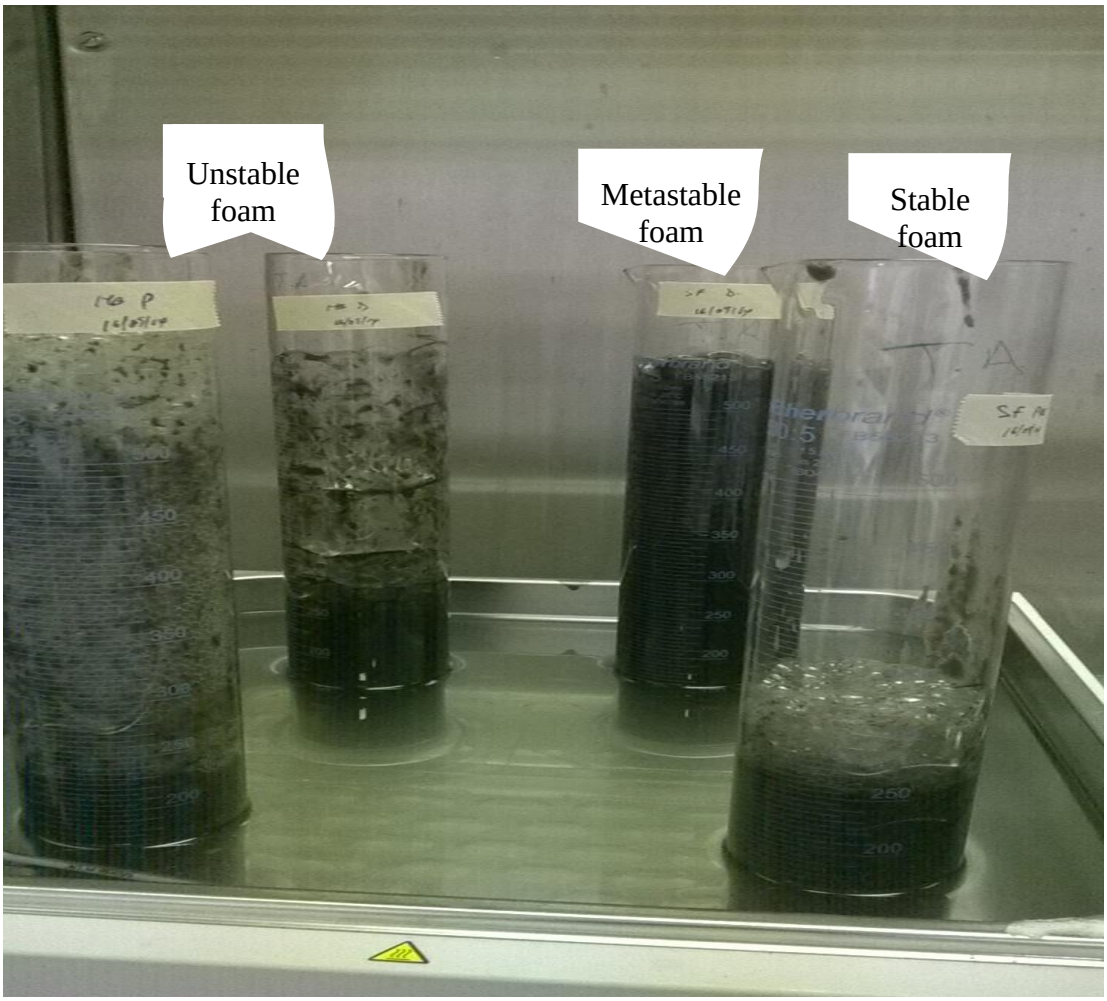


Figure 2-9: Foaming of AD sample in a laboratory experiment at Heriot Watt University induced by Alka-Selzer®

It is also worth noting the difference in characteristics between AD foaming and foaming because of agitation of a mixture of triton x100 in unionised water as shown in figures 2-9 and 2-10 respectively. Although both foams were stable, the AD foam was dark, thick and consist of particles that are left on the wall of the cylinder as it collapses while there is no particle found in the triton x100 foam. This accumulation of particles will impact the AD operation as the particles consist of both biodegradable particles and microorganism. The inability of the microorganism to remain in the digester sludge and aid the digestion process and the unavailability of the biodegradable material to be acted upon reduces the efficiency of the digester In general when foaming occurs in AD, it tends to reduce the production of gas by up to 40% (Moeller, et al., 2010).

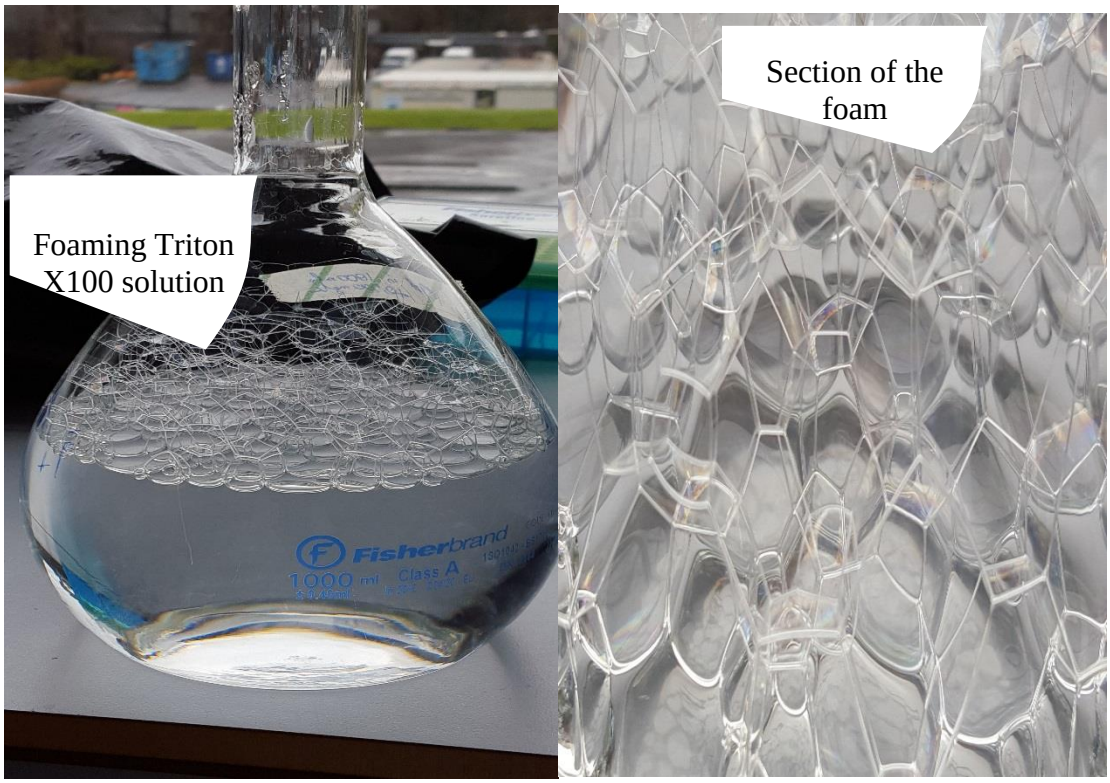


Figure 2-10: Foaming from mixing Triton X100 with unionised water

2.6.3 Why foaming in anaerobic digesters?

Various research studies have been carried out to enable a better understanding of foaming in AD. In a survey and laboratory study carried out by van Niekerk et al. (1987), high ratio of waste activated sludge (WAS) to primary sludge (PS) was identified as one of the causes of foaming in AD. On the contrary a laboratory scale investigation of foaming in AD conducted by Ross and Ellis (1992) reported thickest foam in digesters receiving a low WAS to PS ration. They related foaming to high organic loading ratio (ORL), hydraulic retention time (HRT), high total volatile acid: total alkalinity ratios (tVA: TA), low volatile solids reductions and low pH.

Nevertheless, Barber (2014) has emphasised that different sludge samples respond differently to further processing when subjected to AD with primary sludge (PS) digesting much better than secondary sludge (WAS) as shown in table 2-1. This reveals differences in composition which may be affecting the process differently, thus it may be worth investigating if this is contributing to the foaming episode and the extent.

Using full scale anaerobic digesters that treat a mixture of PS and thickened waste activated sludge (TWAS), Pagilla et al. (1997) investigated causes and effects of foaming and concluded that gas-mixed digesters are more susceptible to foaming than mechanically-mixed digesters with an increase in foaming occurring at extreme levels of Nocardia filaments. A further research by Westlund et al. (1998) reported Microthrix filamentous bacteria as the cause of foaming in a full-scale AD in Stockholm. Though researchers believe that increase in advanced biological treatment processes and the corresponding changes in the quantity and characteristics of the waste activated sludge are the key factors, other aspects such as temperature changes, insufficient mixing, shock loads and hydrophobic substances has been identified as causes of AD foaming (Barber, 2005; Barjebruch et al., 2000). Table 2-5 shows a summary of various causes of foaming as identified in literature:

Table 2-5: Summary of various causes of foam in AD

Sludge feed characteristics	Digestion process related characteristics	Digester operating conditions	Digester configuration
•Surface active agents in feed [Proteins, Lipids, Fat, Oil and Grease(FOG), Detergents] •Foam causing filaments in feed (Particularly Microthrix Parvicella and Gordon Amarae) •Grit/Inert materials	•Organic Loading (Overloading and Inconsistent loading •Volatile Fatty Acid productio (VFA) •Gas production	•Temperature •pH •Mixing	•Digester shape •Sludge withdrawal and gas piping

With the complexity of the foaming mechanism, recent research has been geared towards differentiating the fundamental cause of foaming from other causes that contribute/enhance the foaming episode. Nafiska G. (2009) found foaming initiators as organic loading rate and surface-active compounds. He discovered these foam initiators to be necessary at the beginning of batch digestion in order to set off foaming but did not have to be present thereafter for regular / daily foaming to be present. A full case study at Lemvig biogas plant by Kouglas et al (2014) indicated the feedstock composition and mixing

pattern of the reactor as important factors to be considered to avert foaming incidence. Subramanian and Pagilla (2014) in their research using full scale cylindrical digesters identified the primary causes of foaming as the presence of filamentous bacteria (*Gordona amarae* and *Microthrix parvicella*). Though some distinct clarification has been tried at separating the AD causes from the contributing factors, nevertheless, solutions to AD foaming cannot be achieved without potentially relating the causes to the contributors. Most of the factors affecting anaerobic digester foaming have been discussed in section 2.5, however, surfactant concentration, soluble protein and filamentous bacteria will be discussed here.

2.6.4 Surfactants

Most literature has emphasised a link between the foaming in AD process and the presence of surfactants, biosurfactants and mycolic acid containing organism (Heard et al., 2008 in Ganidi, et al., 2009). The surfactants and biosurfactants are assumed to initiate ASP foaming though critical concentrations for foam initiation have not been quantified due to numerous compounds involved and their variability between different sludge (Heard et al., 2008 in Ganidi, et al., 2009).

Surfactants enter waters and wastewaters mainly by discharge of aqueous wastes from household and industrial laundering and other cleansing operations. The surface-active agents are broken down to simpler compounds (organic acids) during anaerobic digestion and are utilised by bacteria and their impact on the foaming potential is unclear (Ganidi, et al., 2009). However, the removal of detergents during AD, especially for the anionic detergents along with their properties as surface active agents results to increased surface activities in sludge that could potentially contribute to foaming events in AD (Ganidi, et al., 2009).

Biosurfactants has a structure that includes a hydrophilic moiety made up of amino acids or peptides, anions, or cations, or mono-, di-, or poly-sacharides. The hydrophobic portion is often made up of saturated, unsaturated, or hydroxylated fatty acids or composed of amphophilic or hydrophobic peptides (Georgiou et al., 1992). This structure enables it to reduce surface tension,

critical micelle concentration (CMC) and interfacial tension thereby making foaming possible. However, biosurfactants have been observed in AD under non-foaming conditions. It is not clear whether an upset in the metabolic activity of micro-organisms in AD resulting in higher production of biosurfactants that facilitates foaming. Due to lack of experimental evidence Ganidi et al. (2009) proposed that there could be no conclusion as to the contribution of biosurfactants to AD foaming.

2.6.5 Soluble protein

Protein plays a major role in the formation of foam. They enter the digestion process as substrates thus are present in the fermentation liquid from the outset either as microbial products or in the form of extra cellular polymers that are bound to solids. During the degradation of protein, ammonium is produced which can have an inhibiting effect in the biogas production thus facilitating foam formation (Moeller, et al., 2010). Ammonium is usually in dissociation equilibrium with ammonia which is a strong cell toxin. The shift of equilibrium in favour of ammonia depends on the increase of the temperature and or the pH value amongst other factors (Moeller, et al., 2010).

Although ammonium ion and total Kajeldah nitrogen (TKN) have been widely investigated and evaluated on its effect on biogas production, ammonia nitrogen has a higher toxicity and greater effect on AD as it can easily penetrate cells and disturb the metabolism of microorganism (Li, et al., 2013). Some studies have suggested that ammonia nitrogen concentrations above 200- 400 mg/l can result in methane production inhibition while some has stated that ammonia nitrogen concentration lower than 1.1g/l or 1.45 g/l in the anaerobic digestion could result in low methane yield and loss of acetoclastic methanogenesis. Thus, the actual inhibiting threshold ammonia concentration can be a function of differences in substrates and inocula, environmental conditions and acclimation periods (Li, et al., 2013).

However, ammonia increase in the ideal range can aid to reach stability in anaerobic digestion process by making the digestion process more resistant to instability and minimise the risk of ammonium Nitrogen limitations for the methanogens (Li, et al., 2013).

2.6.6 Filamentous microorganism

These microorganisms consist of bacteria, fungi and algae whose cells do not detach from one another after cell division as they are surrounded by a sheath. Over 30 different species have been observed in activated sludge mainly consisting of bacteria; most of them do not have a name rather a number as their features are not yet known completely (Eikelboom, 2000).

Filamentous species are normal microorganism present in activated sludge contributing to the treatment process that can either be free or bound to flocs. However, A huge presence of filamentous microorganism result in bulking sludge thereby reducing the settling and dewatering features of the sludge and subsequently leading to scum formation in AD process (Mesquita, et al., 2013).

Although researchers have linked filamentous microorganism to foaming in AD, it is obvious that the presence of filamentous microorganism in the activated sludge process is a prerogative to having an impact in AD. This is because most filamentous microorganism such as *Gordona amarae* and *Microthrix parvicella* that have been linked with AD foaming are aerobic microorganism that will not naturally proliferate in anaerobic conditions. Hence AD treating sewage sludge that do not have the activated sludge treatment plant experiencing bulking will not likely have a challenge of foaming arising from filamentous microorganism.

The abundance of filamentous microorganism in activated sludge is often measured using filamentous index (FI) on a scale of 0-5 (from none to ubiquitous) with a difference of approximately a factor of 10 between the successive FI classes (Eikelboom, 2000). FI is usually determined by comparing image of sludge at low magnification with a series of reference photograph of the various FI classes as shown in figure 2-11.

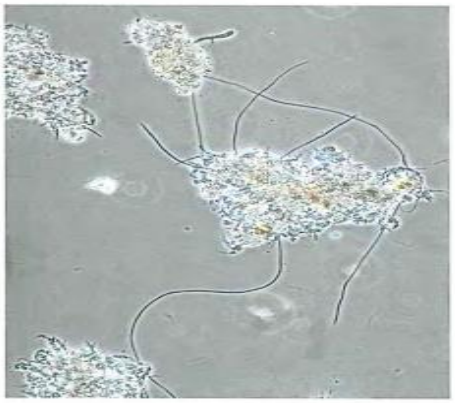


Figure 31 FI = 2; robust filaments (150×).

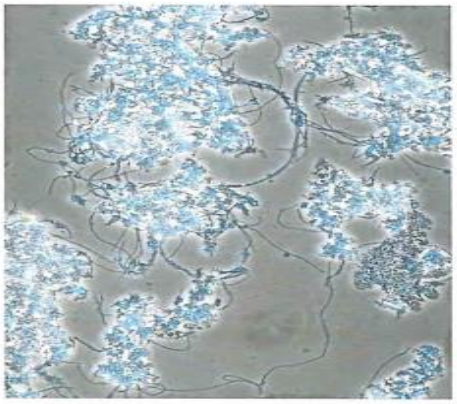


Figure 32 FI = 2; thin filaments (300×).

Figure 2-11: Typical reference photograph for visually determining the FI of Gram stained sludge. (Eikelboom D.H, 2000)

The effect of the filament on the settling velocity of the sludge is slight but become more significant when it is above 3 especially when there is presence of robust filaments leading to bulking sludge (Eikelboom, 2000). Gram stain is often used to make the filament more visible as filaments hidden in flocs cannot be properly observed. Gram negative and Gram positive bacteria stain red and blue respectively. This is as a result of the composition of the cell wall of the micororganism (Eikelboom, 2000). Gram positive bacteria make up the greater percentage of sludge found in high loaded plant while low loaded plant consist mainly of Gram negative. The result of the staining could also depend on the age of the cell for some species as young cells stain red and older ones stain blue resulting in two colours in one filament. Fungi and protozoa do not stain evenly or not at all with Gram stain (Eikelboom, 2000).

Due to the subjective nature of identifying the microorganism when using gram stain, it has been identified by Eikelboom (2000) as a very quick and useful method of understanding the operating condition of the activated sludge process but not very accurate in specifying the exact micoroganism.

2.7 The AD Microbiome

Since the activities of the microorganism form the basis of anaerobic digestion process, an in-depth knowledge and understanding of these microbial communities could be of great use in proffering solution to

anaerobic digester perturbation in order to enhance performance. Microbial analysis of AD has been focused on methanogens as they are the most adversely affected microorganism in anaerobic digestion process (Lee, et al., 2014). However, there has not been any report comparing the microbial composition in AD before, during and after foaming occurrence. Thus, the relationship between the changes in microbial communities during AD operation and the role they play in AD perturbation are not implicit.

2.7.1 Methods for microbial analysis

Understanding AD operation as well as any related problem occurring in the system requires sludge characterisation with respect to biomass content. This is because the microbial features including their internal structure, chemical composition and ecology determine the transport properties and chemical reaction rates of the AD system (Chu & Lee, 2004). Observation of microorganism could be traced back to seventeenth century when Antonij van Leeuwenhoek declared the presence of “little animals” existing in water. He proved this concept by building the first microscopes capable of viewing single-celled organisms in the late seventeenth century (Dobell, 1932).

In the field of wastewater, the earliest research on microbial composition of AS was focused on visualisation of the physical nature of aggregated biomass, the type/abundance of filamentous microorganism and consequently its relation to settling and compaction (Mesquita, et al., 2013; Shah, et al., 2011). Further developments in the nineteenth century ranging from bright field, phase contrast or fluorescence microscopy has led to the ability to determine the AD biomass characteristics in solid-liquid separation.

Consequently, isolation and cultivation for microbial characterization became a common approach after pioneering work carried out by Robert Koch resulted in the discovery of link between disease and causative microbial agents (Ben-David & Davidson, 2014). Yet, only about 1% of microorganisms could be cultured limiting the ability to characterize most microorganisms in environmental samples (Fry, 2000). However, the introduction of ribosomal RNA genes as molecular markers for microbial classification and Sanger sequencing in the late 1970s enabled the identification and classification of

uncultured sample containing various mix of microorganism (Woese & Fox, 1977). Other developments in molecular techniques include the following; fluorescent in-situ hybridization (FISH), polymerase chain reaction (PCR), rRNA genes cloning and sequencing.

Consequent to the discovery of substantial oligotrophic bacterioplankton in the Sargasso Sea in 1990 by Giovannoni *et al*. utilising clone libraries of eubacterial 16S rRNA genes advanced the use of molecular techniques specifically targeting the 16S rRNA gene to investigate environmental samples (Hori, et al., 2006). Currently, nucleic acid based molecular techniques have been mostly used for the analysis of archaeal and bacterial communities existing in AD (Raskin, et al., 1994; Mladenovska, et al., 2003; Kobayashi et al., 2008, Siggins et al., 2011, Shah, et al., 2014)

2.7.1.1 Bright field microscopy

Bright-field microscopy is prevalently used as a basic tool in microscopy because it is easy to set up with only basic equipment requirement such as a standard light microscopy. The conventional microscope makes use of visible light of 400-700 nanometers to illume and produce a magnified image of a sample. The extent of magnification is limited by the wavelength of visible light necessitating adjustment in the source of light through the iris diaphragm. However, using special immersion oil placed on a glass cover over the specimen could yield very good result as the immersion oil has the same refraction as the glass and improves the resolution of the observed specimen (Eikelboom, 2000).

In addition, applying some staining methods such as Gram, Methylene Blue, Neisser and Sudan Black B have proved very useful technique in highlighting features of microorganisms when placed under bright field microscopy. Filamentous bacteria were identified according to Eikelboom (2000). Ganidi et al., (2011) used Gram Safranin and Neisser stains for staining of filaments in AD sludge. Observing them under Light microscopy, they identified five filament species in the full scale digester, including *Norcodia limicola I and III, Microthrix Parvicella, 0041, 0581*. The overall abundance of the identified microorganism did not exceed a filament index (FI) of 3 according

to a scale ascending from 1 to 5 in abundance. During the observation period, species *N. limicola I and III, Microthrix, 0041, 0581* recorded maximum FI of <1, 1.5, 3 and 2.5 respectively. It was acknowledged that prior to recording foaming at full scale, the abundance of 0041 species increased from 2 to 3 on the FI scale (Ganidi, et al., 2011).

Other forms of microscopy include; phase-contrast microscopy, fluorescence microscopy, confocal laser scanning microscopy with each having its merit and demerits as explicitly discussed by Mesquita, et al. (2013)

2.7.1.2 Fluorescence In-Situ Hybridisation

Using a fluorescent probe, FISH has regularly been used in the study and identification of specific microbial groups by determining if a particular gene is present in a sample and/or if that gene is being expressed under a given set of conditions. Fluorescent probes are short sequences of DNA labelled with fluorescent dye designed to recognise and attach to specific genetic regions of microbes (16S rRNA sequence in cells) that will differentiate them from other groups. When these probes are applied a fluorescent microscope can be used to detect the presence or absence of individual microbial groups. Fluorescent probes are used with different staining at the same time for varying targets to decide the portion of a population that different microorganism belong to with the help of a different filter and stain known as DAPI. DAPI binds DNA non-specifically resulting to an image showing the combined microbial population.

The key disadvantage of FISH analysis relates to the need of previous knowledge of the microbial population and dominant microorganism under study in order to be detected. In a recent study, Hirakata, et al. (2015), investigated the prokaryotic community structure of the anaerobic ciliate, Metopus sp. using rRNA sequencing, fluorescence in situ hybridization (FISH), and transmission electron microscopy (TEM). The FISH analysis using the oligonucleotide probes Mg1200b and Cla568 demonstrated that these prokaryotes were localized within Metopus cells. These results identify *Methanoregula boonei* and *Clostridium aminobutyricum*-like prokaryotes as novel endosymbionts of *Metopus* ciliates (Hirakata, et al., 2015).

2.7.1.3 Flow-Cytometric Analysis

This is another technique that can be carried out when fluorescent tags are applied to microbial populations. Fluorescent signal is used to count or sort individual genotypes out of groups of cells. Flow cytometry differentiates complex bacterial populations based on fluorescence and size differences among the cells. In combination with a sorting unit, defined Flow cytometry populations can be physically separated and exposed to further taxonomic analysis. Nevertheless, the application of these combined methods is challenging due to large number of sorted cells needed for subsequent DNA-based analysis (Mou, et al., 2005).

2.7.1.4 'Omics' methodology

The field of molecular biology connects with biology and chemistry, and more precisely, genetics and biochemistry. Rapid advances in molecular technologies have paved way for the combination of 'omics' approaches such as metagenomics, metaproteomics, metatranscriptomics and metabolomics to provide invaluable information on the genetic, functional and metabolic activities of microbial communities (Fritz, et al., 2013). Zhang et al., (2010) highlighted the three major features shared by all omics data that differentiates them from traditional methods inter alia; omics method are high-throughput and data-driven; the generation and analysis of the data are carried out for the whole microbial population; large amount of data generated require a good bioinformatics knowledge to discern.

Each 'omics' technique is aimed at diverse constituents of the microbial community, thereby providing distinctive information that proffers a more in-depth knowledge on complexity existing in various ecological unit. However, with a high diversity in phylogenetic and functionality of microorganism present in AD, combining 'omics' (integrated omics) datasets could be very useful in identifying significant players and the different important functions that they undertake in such a complex system. This will require adequate sample preparation, data clean-up and data analysis to permit useful deduction from 'omics' data.

2.7.1.4.1 Metagenomics

Environmental Metagenomics involves the study of organisms that make up microbial population by analysing the DNA existing in it (Handelsman, et al., 1998). Demand for human genome sequencing resulted in a massive increase in availability of comparatively low cost high throughput next generation sequencing (NGS) making it possible to characterise the complexity of microbial population in their natural environment thereby eliminating the need to isolate and culture individual species.

NGS stands out in their ability to produce a greater number of base pairs as many samples could be combined in a sequencing run thus drastically reducing the time required to sequence a complete genome. These advancements have provided valuable information about genomic diversity, composition, function and metabolic capability of a particular environment as well as evolutionary analysis in animals and plants (Gupta & Gupta, 2014).

Broadly, metagenomics analyses involve the basic stages of DNA extraction, library preparation, sequencing and data analysis (Bioinfomatics). The difference between the various next generations sequences depend mostly in the method of sequence interalia; Shotgun metagenomics, whole genome sequencing and 16S rRNA sequencing. Although each of the NGS have different purpose for which they are best used, 16S rRNA seque (Guermazi, et al., 2008)ncing have been widely adopted in most environmental analysis as it permit the study of phylogeny and taxonomy of microorganisms present in samples obtained from environment that are difficult or impossible to study because of their complexity (Garza & Dutilh, 2015).

Thus, 16S rRNA sequencing provides an approximation of the percentage of the population that has not been cultured enabling researcher to identify areas for further study (Rappé & Giovannoni, 2003). In 16S rRNA, high-throughput sequencing of environmental samples has been achieved by extraction, followed by purification of DNA and RNA also known as reverse transcribed into cDNA and subsequent amplification of targeted genes with universal primers. The universal primers can be targeted either on the highly conserved or hypervariable regions.

The highly conserved regions are applied on the amplification of the whole gene while the hyper-variable regions are linked with specie specific sequencing that can discriminate between different bacteria and archea. With each primer set barcoded with short oligonucleotide tags and sequencing adapters, the pooling of multiple samples together for sequencing could be achieved. The data collected from the sequencing process will be filtered and subjected to bioinformatic analysis by means of taxonomic tools such as QIIME (Caparaso, et al., 2012).

These bioinformatics tool have been very useful for analysis such as; quality control, assembly, gene detection, gene annotation, taxonomic analysis, comparative analysis, and interpretation. In addition, there exist, many databases of gene and bioinformatics software for metagenomic data analysis (Abbasian, et al., 2015).

Metagenomics analysis was carried out on microbial communities from a full-scale AD fed with a mixture of maize silage, green rye and chicken manure (Jaenicke, et al., 2011). The most abundant phyla for degradation of the mixed substrates were Firmicutes and Bacteroidetes while Clostridia were perceived to be responsible for syntrophic association with hydrogenotrophic methanogens. Since, methanogens are mostly affected by AD perturbation, most studies were dedicated to methanogenic composition and how it influences biogas production (St-Pierre & Wright, 2014)

Though 16S rRNA sequencing data depict a wider range of microbiome diversity with a lower sensitivity and resolution due to bias by unequal amplification of species, whole metagenomics sequencing tends to be less efficient in determining rare species (Cox, et al., 2013). Short gun sequencing is more efficient for small number of low complexity sample however, the challenge lies on limitation in the length of data and rigorous workflows (Shah, et al., 2011). Thus, the ultimate decision on the method to adopt is highly dependent on the objective of the research project.

2.7.1.4.2 Metatranscriptomics

Metagenomic provides a snapshot of gene content and diversity in a microbial community. These genes are made up of deoxyribonucleic acid (DNA) which is a long, winding molecule that encompasses the instruction necessary for structuring and preserving cells. DNA exists in the form of base pairs of guanine/cytosine and Adeline/Thymine supported by strands of phosphate and sugar organized to form about 20,000 to 25,000 genes. In order to execute the instructions, DNA must be "read" and transcribed by copying the DNA into RNA (ribonucleic acid). These gene readouts are termed transcripts while an assembly of all the transcripts existing in a cell is known as transcriptome. The major type of RNA that plays a crucial part in protein production is the messenger RNA (mRNA). This is achieved by delivering mRNA transcripts to molecular machines located in the cell's cytoplasm termed ribosomes; subsequently, the sequence of chemical letters in the mRNA is read by the ribosomes followed by assembly of amino acids into proteins.

Metatranscriptomics emerged as a unique process that involves recovering RNA from environmental sample which has gone through mRNA fortification and cDNA generation, and subsequently sequencing it, to determine microbial activities by measuring in-situ gene expression (Zarraonaindia, et al., 2013). Metatranscriptomics is based on the principle that although virtually every cell contains the same genes, different cells show different patterns of gene expression. These dissimilarities are responsible for the varying features and characteristics of different cells and tissues.

Since an RNA sequence reflects the DNA sequence it was transcribed from, analysing the complete assembly of RNA help researchers to infer when and where each gene is turned on or off in the cells and tissues of an organism. This knowledge deepens our understanding of what constitutes a specific cell type, how that type of cell normally functions, and how changes in the normal level of gene activity may reflect or contribute to perturbation in an environmental system (Gracey, 2007; Evans & Hofmann, 2012)

Metatranscriptomics can be performed by micro array or sequencing. With microarray, the expression profiles of treated and untreated cell cultures,

normal and tumour tissues developmental stages of an organism or tissue, and different tissues can be compared to derive more knowledge of the transcriptome. Using the more recently NGS, whole transcriptome sequencing could be performed to identify and quantify novel and known transcripts while targeted transcriptome sequencing could be applied for simple gene level expression analysis.

NGS can detect subtle changes in expression level in a hypothesis-neutral environment thereby providing an understanding of biological response to environmental changes. Some tools for the assembly of metatranscriptomic reads have been developed such as a reliable pipeline that can process Illumina-RNA-seq metatranscriptome data by linking the sequence reads to reference gene databases, their assigned functions, and predicted phylogenetic origin (Leimena, et al., 2013).

Zakrzewskiet al., (2012) reported that firmicutes were dominant in a study analysing sludge samples from an AD treating maize silage, green rye and chicken manure. Transcripts revealing enzymes for methanogenesis were mostly represented with taxonomic classification belonging mostly to Methanomicrobiales (Zakrzewski, et al., 2012)

Though transcriptomics could efficiently identify transcribed genes, however, its inability to account for post transcriptional regulation creates concern as mRNAs that could not translate to protein are not reported. In addition, due to the short half-life of the mRNA content, the task of mRNA enrichment could be onerous and challenging thus requiring that samples should be carefully handled and stored.

2.7.1.4.3 Proteomics

Proteomics analysis enable a comprehensive identification of proteins expressed in a given microbial community thus widening knowledge on microbial activities Though studies relating to proteomics had been carried out as far back as in 1975 with the introduction of two dimensional gel (Scheele, 1975; Klose, 1975; O'Farrell, 1975), the word proteomics was first used in 1995 as a study of total protein content of a cell, tissue or organism

(Anderson & Anderson, 1996). In these previous studies, protein could be separated and visualised, but they could not be identified (Graves & Haystead, 2002). Since then, sequencing of protein has evolved through Edman degradation (Edman, 1949), microsequencing technique for electroblotted proteins (Aebersold, et al., 1986) and finally mass spectrometry technique. Mass spectrometry has been widely accepted as it is more sensitive, tolerant of different protein mixtures and works well with high throughput operations (Graves & Haystead, 2002).

The proteome of a cell reflects the immediate environment as protein can vary from posttranslational modification, undergo translocations within a cell, or synthesized (Graves & Haystead, 2002). Thus, efficient sampling and extraction techniques are essential to generating good proteomic result. The different forms of proteomics include protein expression, structural and functional. Carrying out a typical proteomic analysis will involve; collection of sample, separation/isolation of protein form a cell/tissue/organism, Mass spectrometry analysis and extraction of structural information of the protein for identification and characterisation using available metagenomics database.

Metaproteomics study was used to assess the microbial functions carried out in situ for a laboratory AD (Gunningle, et al., 2015). It was observed that variation in digester temperature resulted in the differential expression of proteins linked with methanogenesis. Bacteroidetes and Firmicutes increased at lower temperature, δ-Proteobacteria decreased in abundance at lower temperature while *Methanomicrobiales* abundance increased at 15°C. *Methanobacteriales* and *Methanosaeta* were reported to be predominant at all temperatures investigated (37°C, 15°C and 7°C). A fascinating discovery was made by Gunuigle et al., (2015) stating that *Methanosaeta* could carry out methanogenesis from carbon dioxide contrary to the idea that it could only produce methane from acetate.

Since protein identification is dependent on a comprehensive metagenomics data (Muth, et al., 2013) which makes it difficult to easily relate protein abundances to microbial activities, proteomics majorly complements other

omics analyses (Tyers & Mann, 2003). Unfortunately, the identification of lots of proteins in a given sample tend to create more questions rather than actually providing a solution (Graves & Haystead, 2002). Thus, it is very vital to have a clear definitive purpose for proteomics analysis such as a specific biological question before proceeding with proteomics analysis.

2.7.1.4.4 Metabolomics

Metabolomics is the analysis of various metabolome which is the small molecule contained in a biological system. Metabolites form the basis for biochemical reactions occurring in living systems such as conversion of glucose into energy, sending and receiving information in the form of chemical signals and provide the foundation for complex macromolecules. The challenges encountered in comprehensive analysis of metabolites is related to the fact that it encompasses various low molecular weight structures which could be Lipids, peptides, amino acids, thiols, carbohydrates, etc. (Zhang, et al., 2012)

Nevertheless, metabolomics offers the most sensitive indicator of phenotype in a microbial community thereby enabling a more accurate inference of vital metabolic processes (Roume, et al., 2013). This is obtainable because, metabolic fluxes are not regulated by gene expression or posttranslational modification compared to metaproteomics and metatranscriptomics (Kim, et al., 2009; Hettich, et al., 2013).

Earlier, detection of metabolites is usually not easily differentiated thereby making it difficult for any single analytical platform to identify all metabolites present in a biological sample. However, the evolution of metabolomics study permits the analysis of large portions of the complete metabolome at once (Zhang, et al., 2012) by combining rapid developments in modern instrumental analytical approaches. Thus, metabolomics goes beyond providing information for specific metabolites of interest to proffering an enhanced knowledge on the complex biochemical changes and interactions taking place in a microbial population.

The instrument analytical platforms include; Gas chromatography (GC), High performance liquid chromatography (HPLC), Ultra performance liquid chromatography (UPLC), Capillary electrophoresis (CE) coupled to MS (Mass Spectrometry) and Nuclear Magnetic resonance (NMR) spectroscopy which are designed to enable separation, detection, characterization and quantification of such metabolites and related metabolic pathways (Zhang, et al., 2003). NMR spectroscopy is mostly suitable for the analysis of targeted metabolites while GC-MS are best suited for volatile organic compounds/derivatives primary metabolites and LC-MS are commonly used as they can avoid chemical derivatization (Zhang, et al., 2003).

MS-based method tends to characterize the entire metabolome (Oresic, 2009). Generally, MS-based metabolomics ensures high selectivity and sensitivity for the identification and quantification of metabolites, although some step in the sample preparation could lead to metabolite loss (Kim, et al., 2009). As a result, applying several techniques simultaneously could enhance the study of metabolites such as GC-MS, LC-MS or NMR. A detailed review of the procedure for the different analytical techniques could be found in Zhang, et al. (2012) and Cevallos-Cevallos, et al. (2009).

Metabolomics have been extensively used in the area of human disease such as studies on gut metabolome, (Le Gall, et al., 2011; Heinken, et al., 2014) human nutrition ((Wishart, 2008), plant analysis (Hall, et al., 2008), drug discovery (Wishart, 2008), etc. Through metabolomics, changes in polyphenolic compounds during berries breeding have been characterized (Stewart, et al., 2007). Combining metabololomics and metaproteomics an investigation on Euglena mutabilis in an acid mine drainage (AMD) system was executed. It was discovered that the protist was performing photosynthetic metabolism thereby secreting organic matter in the environment selectively (Halter et al., 2012). In a review by Kushalappa et al., (2008), they were able to ascertain the potential for detecting and understanding food spoilage using food spoilage post-harvest metabolomic analysis.

Unfortunately, the difficulty involved in developing a comprehensive metabolomics of a complex microbial sample with abundance of metabolites has limited the application of metabolomics in AD study. The existence of this gap, highlights the need for further study of AD samples using metabolomics analysis.

2.7.1.5 Integrated "Omics"

A variation in the combination of omics have been applied in recent microbial analysis to develop a better understanding of the microbial community in various research such as; biological wastewater treatment (Narayanasamy, et al., 2015), acid mine drainage biofilms (Denef et al., 2010), plant system biology (Fukushima, et al., 2009), a lake ecosystem (Lauro *et al*., 2011), and intestinal microbiota (Kolmeder *et al*., 2012).

Using metagenomic and metaproteomic analysis on AD batch fermentations for a mixed substrate of maize, straw and hay enabled the identification of microorganisms involved in plant carbohydrate degradation (Hanreich, et al., 2013). Proteins were predominantly assigned to the taxa *Clostridiales* and *Bacteroidales*. *Clostridiales* were key to the degradation of cellulose and decreased while *Bacteriodales* increased during the fermentation process. (Hanreich, et al., 2013). It was further identified that although methanogenic organisms were a minor group in the community yet there was expression of many methanogenic enzymes (Hanreich, et al., 2013)

Integrating 'omics' datasets enhances the understanding of microbial population and their activities in complex systems such as AD. This offers an adequate guide to system optimisation. NGS can also be used to characterise the process at start-up or during changes in operating conditions or feedstock. This is vital for a better understanding of process dynamics under critical operational phases.

2.8 Modelling anaerobic digester process

Modelling has been a very useful tool that describes a system operation to enable an efficient understanding of how it works to predict its behaviour. Thus, modelling anaerobic digestion process plays a vital role in efficient

development of design process and implementation of operational control that enhance the system performance.

Proficient application of model simulation eliminates the need for continuous experimentation with greater reliance on online assessment and adjustment of operating variables (Olsson, et al., 2005). The resultant effect within the context of the AD is a more stable system, greater solid destruction, improved biogas quality with a higher methane content leading to a reduction in overall running cost. Adequate model structures should be simple, represent the most relevant cause-effect interactions and be able to identify the values of the unknown parameter for the present and future data.

Structurally, models can either be empirical or mechanistic (Osson & Newell, 2001). Mechanistic models assume that a complex system can be understood by examining the workings of individual process and the way they relate through mathematical and kinetic expressions. Empirical models on the other hand are constructed based on direct observation, measurement and extensive data records. Alternative classification of models based on the type of knowledge via which it was developed include (Khan, 2012):

White Box: This type of model is also known as the mechanistic model and the model development is based on detailed knowledge and structure of the system to be modelled. The knowledge enables the development of model equations from general balance equations on mass and other conserved quantities to generate a set of differential equations

Black Box: It is a modelling technique that does not require the model developer to have prior knowledge of the internal mechanism and working of the system. It only examines the fundamental aspects of the system through collected data. A ‘black-box’ also known as an input-output model is mainly driven by the measured data obtained from the process thus are not good extrapolators as they are only as good as the data that were used to calibrate them (Kythreotou, et al., 2014). This type of model is also known as empirical model.

Grey Box (White box + Black box = Grey box): It is a modelling technique that is based on a combination of limited knowledge of the internal working of the system and some knowledge of fundamental aspects of the system. Typical example is the neuro-fuzzy model to be discussed later in this chapter.

With a clear understanding of the different forms of modelling, the following factors need to be considered to ensure that the model to be developed is a true representative of the system being modelled.

2.8.1 Data collection

Successive observations of a system are not necessarily exact resulting in variability in data collected (Montgomery & Runger, 2011). Thus, equation 2-5 is a convenient representation of the impact of variability on collected data.

$$A = ħ + Ɛ \qquad \text{Equation 2-5}$$

Where

A= Observation

ħ = constant

Ɛ = random disturbance

The constant remains the same with every measurement but small changes in the environment, variation in test equipment, difference in the individual components, etc. will vary the value for Ɛ. These factors yielding the random disturbance is seldom absent resulting in stochastic systems. It is vital that adequate consideration is given to the presence and impact of these random disturbance during data collection by ensuring that at least triplicate sample is taken for every observation such that the average calculate wil be a fair representative of the required data.

The following data collection methods have been used for developing engineering models: retrospective study using historical data; observational study by recording process data while disturbing the system as little as possible; designed experiment by making purposeful change in the controllable variables of the system/process and assessing the resulting data from the system to ascertain the variable that is responsible for the changes observed in the output data (Montgomery & Runger, 2011). Since designed experiment is the most representative of the various methods of data collection, further discussion on incorporating random disturbance in the data collected will be focussed on designed experiment.

2.8.1.1 Designed experiment

The validity of the conclusions drawn from data is highly dependent on how the experiment was carried out and data collected. Statistically based experimental design are very crucial to discovery of new basic phenomena, commercialisation of new technology (product/process development) and improvement of existing products/processes (Montgomery & Runger, 2011). It is best to employ designed experiment sequentially. This involve identifying the controllable variables, carrying out screening experiment to determine which variables are most important, and follow-up experiment to determine the necessary adjustment to the control variables that will result in process improvement.

Experimental design could be single-factored or multi-factored. A single-factored experiment consists of one control variable altered at different level. The best approach to statistical design of single-factored experiment is to ensure randomisation of the replicates such that the effect of any nuisance is approximately spread out. For multi-factored experiment (factorial experiment), each complete trial or replicate of the experiment is carried out with all possible combination of the levels of the factors to be investigated (Montgomery & Runger, 2011).

Nevertheless, it is crucial to maintain a balance between statistical accuracy of the experimental design, feasibility, and economic viability of the project (Montgomery, 1984). During the experiment, focus should be on

randomisation, measurement accuracy and maintaining uniformity of experimental environment as efficiently as possible (Montgomery & Runger, 2011). Once the experiment is completed, the data collected are statistically validated prior to feature selection.

2.8.1.2 Stochastic process and synthetic data generation

Stochastic processes are systems which change with time according to certain probabilities. The most basic example is random walks. Stochastic process could be temporal, spatial or spatio-temporal. Temporal processes involve variation in time interval. If the time points are specific, then it is known as discrete-time process while if the time point varies infinitely throughout the real line then it is a continuous-time process. Spatial stochastic process represent location in space rather than time for different random vectors (set of various time intervals). In such scenario, stochastic process could be the generalisation of random vectors. When a process involves time and space, it will result to a spatio-temporal stochastic process. In addition, stochastic process could be stationary or non-stationary.

Stochastic process modelling has been important component in the world of energy system and environmental analysis especially in forecasting and generating synthetic data. Due to nature and complexity of stochastic process of which AD foaming is a typical example, obtaining comprehensive data set for such modelling could be challenging hence the need to develop synthetic data to offset any limitation in available data without losing the basic dynamics of the system.

Synthetic data generation is the technique of applying specific algorithms that are designed to create representative data about a system/process, thus, direct measurement does not obtain them (Deng, 2002). Selecting the most appropriate algorithm for synthetic data generation can be challenging since different algorithms are more suitable to diverse testing datasets based on specific factors such as data distribution and dimension. Synthetic data has proved very useful in developing more efficient and robust model in various sectors by creating large volumes of data for performance and load testing, reducing infrastructure by covering all combinations in the optimal minimum

set of test data, improving existing subsets of production data with rich, sophisticated sets of synthetic data, and recently, eliminating the risk of data breach by creating production-like data without sensitive content.

Traditionally, synthetic data have been used extensively in fields such as; water resource management/Climate change, data mining and energy systems. The focus in this study will be the fundamental data generation methods applied in these systems. Since these are real world scenario, filled with lot of stochastic process, stochastic model of synthetic data generation tends to be readily adaptable as it factors the random disturbance. McMahon & Adeloye (2005) identified that streamflow is a function of time and space, fluctuates in a random manner and cannot be predicted with certainty thus they emphasised the need to generate alternative sequences of streamflow that possess the same statistical properties as the historical data using stochastic stream flow models. They argued that the sample could be applied in parameterising and constructing theoretical probability distributions used for further analysis of the streamflow that would not have been possible using a single historic record.

Stochastic models could be parametric or non-parametric (McMahon & Adeloye, 2005). Parametric stochastic models mainly consist of linear mathematical formulation in which the probability distribution of the stochastic process must be specified explicitly, and the parameters of the model estimated using available historical data. Thus, the accuracy of the data affects the usefulness of the generated data. Parametric models include: auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA), broken line model and fractional Gaussian model. Typical example of ARMA model that have been extensively used in hydrology models is Markov lag-one auto-regressive model AR (1). Non-parametric models involve re-sampling methods thus does not fit theoretical probability distribution in describing the stochastic process. Thus, non-parametric models tend to generate more representative synthetic data compared to the parametric models as parameter estimation uncertainty does not influence them. Examples of non-parametric methods include moving window and formalised bootstraps.

Given the time to be taken for the experiments associated with the investigation, it will not be possible to collect sufficient volume of data for driving the empirical models of foaming to be developed. Thus, stochastic data generation will be relied upon for extending the data collected during the experiments. The parameter of the stochastic model will be estimated using the collected data thus ensuring that the extended data to be produced will have the same basic parameters as the experiment data used for its calibration.

McMahon & Adeloye (2005), identified that simple stochastic models such as AR (1) suffice for the generation of stream flows for reservoir simulation as developing other complex model could be challenging and require higher volume of data to achieve a positive result. The AR (1) model is fully specified using equation 2-6.

$$Q_{i+1} = \mu + \rho(Q_i - \mu) + Z_i \sigma\sqrt{(1 - (\rho_i)^2)}$$ Equation 2-6

Where;

i =1, 2, . . . , N, and, N is the number of time periods

Q_i = values for the variable at ith time period

Q_{i+1} = $(i + 1)$th time period

μ = mean of the observed data)

ρ = lag one serial coefficient

Z_i = standardising normal variate, i.e. with a mean of zero and variance of 1

σ = standard deviation of the observed data

The model in equation 2-6 has 3 parameters, namely the mean (μ), the standard deviation (σ) and the lag-1 serial correlation coefficient (ρ). These

three parameters must be estimated using equations 2-7a, 2-7b and 2-7c respectively.

$$\mu = \frac{1}{n}\sum_{i=1}^{n}(Q_i)$$

Equation 2-7a

$$\sigma = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(Q_i - \mu)^2}{n-1}}$$

Equation 2-7b

$$\rho_k = \frac{\frac{1}{n-k}\Sigma^{n-k} Q_i Q_{i+k} \; - \frac{1}{(n-k)^2}\,\Sigma^{n-k} Q_i \Sigma^{n-k} Q_{i+k}}{\left[\frac{1}{n-k}\Sigma^{n-k} {Q_i}^2 - \frac{1}{(n-k)^2}\,(\Sigma^{n-k} Q_i)^2\right]^{0.5} \left[\frac{1}{n-k}\Sigma^{n-k} Q_{i+k}^2 \; - \frac{1}{(n-k)^2}\,(\Sigma^{n-k} Q_{i+k})^2\right]^{0.5}}$$

Equation 2-7c

The variable Z_i confers a probability distribution on the data. As noted above, if Z_i has a norminal distribution, then the data is also normally distributed.

While it might not be wholly wrong to use the normality assumption for most data series, there are situations where such an assumption could be wrong. McMahon and Adeloye (2005) provide alternative distribution hypothesis that could be used in such cases e.g. the log-normal or the gamma. It is also possible to transform the data to normal using e.g the Box-Cox transformation (Soundharajan, et al., 2016).

2.8.1.3 Variable selection

Variable selection is the identification of those variables that contribute to the modelling task and omit others that are not relevant. Variable selection is important for model learning as it help to resolve the curse of dimensionality

i.e. the situation where the number of data samples required to estimate some arbitrary multivariate distribution increases exponentially as the number of variables increases in historical data analysis (Powell, 2007).

However, in experimental design, the key to adequate variable selection is dependent on practical experience, prior academic knowledge and use of physical theory to explain some of the relationships existing factors and responses about the system under survey. This type of non-statistical knowledge is vital to choosing factors, determining factor levels, deciding on the number of replicates, interpreting the results and making useful deductions in conjunction with the result of the statistical analysis (Montgomery, 1984). Thus, the use of statistical analysis does not waive the need for engaging once expert knowledge of the system in proffering solution to the challenge.

The most common approach to evaluating significant variables involves establishing the correlation between a specific input variable and the desired output variable (Guyon & Elisseef, 2003; Powell, 2007). In other circumstances other methods could be deployed such as: starting the modelling process with minimum number of variables and gradually adding one variable every time or vice versa (eliminating method); algorithmic variable selection based on combinational optimisation (Langley, 1994; Powell et al., 2007); and optimal variable selection that requires an exhaustive search of all possible variables thus requiring a huge number of evaluations of the model accuracy. Dalmau et al., (2008) used a neural network wrapper approach with a hill climbing elimination strategy in selecting the most relevant variables for foaming diagnosis in anaerobic digestion process. Inflow rate, total organic carbon and carbon dioxide percentage were among the most relevant variables.

However, when large historical data are used for developing empirical models, extracting features from variables could be very beneficial by constructing combinations of variables that promote the curse of dimensionality while maintaining the accuracy of the data. Through extracting features from variables, large set of data are simplified by

transforming the measurement space to a lower dimensional feature space with higher manageable level (Rustum & Adeloye, 2011). This method ensures that training data set is a comprehensive representative of the system to be modelled such that similar patterns are not repeated as repeating patterns do not offer any new learning benefit rather it slows down the training process significantly.

The fundamental approach to extracting features from variables is the aggregation of variables contributing essentially the same information in a logical fashion. The result of the transformation process is the generation of fewer high level variables than the observed data thereby increasing the accuracy of the model with less computation demand. (Torkkola, 2003). Extracting features from variables may involve linear or nonlinear combination of the observed data (May and Jain, 1995; Laine, 2003). Principal Component Analysis (PCA) and Kohonen Self-Organising Map are typical examples as well as the most well-known linear and non-linear techniques respectively (Kohonen, 1982).

2.8.1.3.1 Choosing adequate method for variable selection

Since most of the statistical analyses establishing the relationship between collected data are dependent on the assumption of normality, a check of normality is crucial prior to selecting the procedure to be applied. Different method for checking normality of a data include inter alia: Visual methods (Frequency distribution-histogram, stem-and-leaf plot, boxplot, probability plot and quantile plot), Kolmogorov-Smirnov (K-S) test, Shapiro-Wilk test, Anderson-Darling test, Lilliefors corrected K-S test, D'Agostino skewness test, Anscombe-Glynn kurtosis test, D'Agostino-Pearson omnibus test, Cramer-von Mises test and the Jarque-Bera test (Ghasemi & Zahedias, 2012).

Ghasemi and Zahedias (2012) observed that the result from these different methods of analysis could vary when presented with the same data. They identified that lack of symmetry (skewness) and peakedness (Kurtosis) are two major ways that a distribution could deviate from normality hence could be a more efficient test for measuring the normality of data.

The skew value of a normal distribution is zero, usually implying symmetric distribution. A positive skew value indicates that majority of the values lie to the left of the mean and vice versa if it is negative. The kurtosis should be zero for a perfectly normal distribution. Distributions with positive kurtosis are called leptokurtic distribution meaning high peak, and distributions with negative kurtosis are called platykurtic distribution meaning flat-topped curve (Kim, 2013).

However, since most data will not be perfectly normal, researchers have established that normality test using absolute z-score for skewness and kurtosis determined from equations 2-8 and 2-9 respectively; could offer a more robust method for testing the normality of data (Ghasemi & Zahedias, 2012; Kim, 2013).

$$Z_{Skewness} = \frac{Skew\ value}{Standard\ error\ of\ skew\left(\sqrt{\frac{6}{n}}\right)}$$

Equation 2-8

and

$$Z_{SKurtosis} = \frac{Kurtosis\ value}{Standard\ error\ of\ Kurtosis\left(\sqrt{\frac{24}{n}}\right)}$$

Equation 2-9

Where n is the sample size

Ghasemi & Zahedias (2012) proposed that when assessing the normality of data the following should be considered:

1. An absolute value of Z-score for skewness or kurtosis greater than 1.96 or lesser than -1.96 is significant at alpha level < 0.05,

2. An absolute value of Z-score for skewness or kurtosis greater than 2.58 or lesser than -2.58 is significant at alpha level < 0.01,

3. An absolute value of Z-score for skewness or kurtosis greater than 3.29 or lesser than -3.29 is significant at alpha level < 0.001.

4. In small samples, values of Z-score for skewness or Kurtosis greater or lesser than 1.96 are sufficient to establish normality of the data. However, in large samples (200 or more) with small standard errors, this criterion should be changed to ± 2.58 and in very large samples no criterion should be applied (that is, significance tests of skewness and kurtosis should not be used)

Hence, using the right normality assessment is crucial to validating the adequacy of the statistical method applied.

2.8.2 Mechanistic model in anaerobic digestion

Modelling of AD process dates to the early 60's and was prompted by the need to optimize AD process using mathematical models (Husain, 1998). However, considering the high complexity and non-linearity of AD process, the modelling has evolved through different phase. The first phase involved very simple models due to limited knowledge and was centred on determination of maximum theoretical biogas rate and unravelling the limiting step in the AD process. The limiting steps considered were mainly Hydrolysis (Hill &Barth, 1977; Eastman and Ferguson, 1981) and Methane fermentation (Buswell & Mueller, 1952; Hill, 1982).

Experimental investigation, further system analysis and the increase in computing capacity led to the development of much more detailed and complex models in recent years. These complex models reflect the dynamics involved in AD process including the kinetics of microorganism growth (Moser, 1988), variation in substrates and various operating conditions.

Overall, the key highlight of mechanistic modelling in AD is shown in figure 2-12.

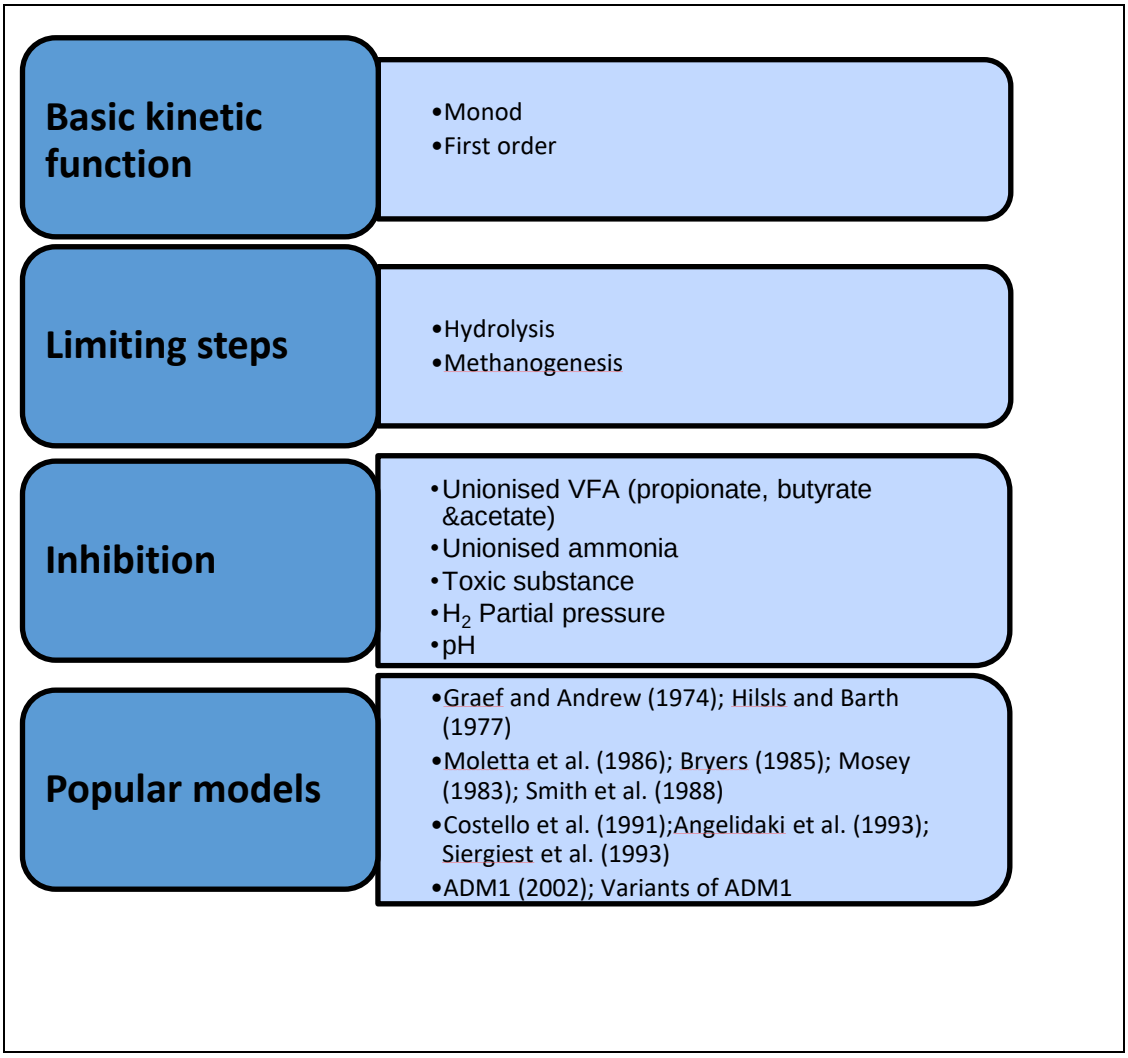


Figure 2-12: Highlights on Mechanistic modelling of AD

2.8.2.1 Bacterial growth kinetics

The basics for mechanistic model of AD have been centred on bacterial growth as the activities of these microorganisms govern the biochemical reactions taking place in the AD which could be the formation (anabolic) or breakdown (catabolic) of cell constituents and metabolites.

Bacterial growths are typically studied using a batch or continuous isolate of the microorganism in a growth medium such that there is full control of the nutrient supply and environmental conditions to achieve optimal growth.

Subsequently, timely measurement of the bacteria increment (number or mass) is recorded and used in developing the growth curve. As depicted in figure 2-13, there are distinct growth phases that can be observed within a growth curve; lag, exponential/log phase, stationary, and death. Each of the phase highlight a specific growth pattern due to physiological changes taking place in the cell culture.

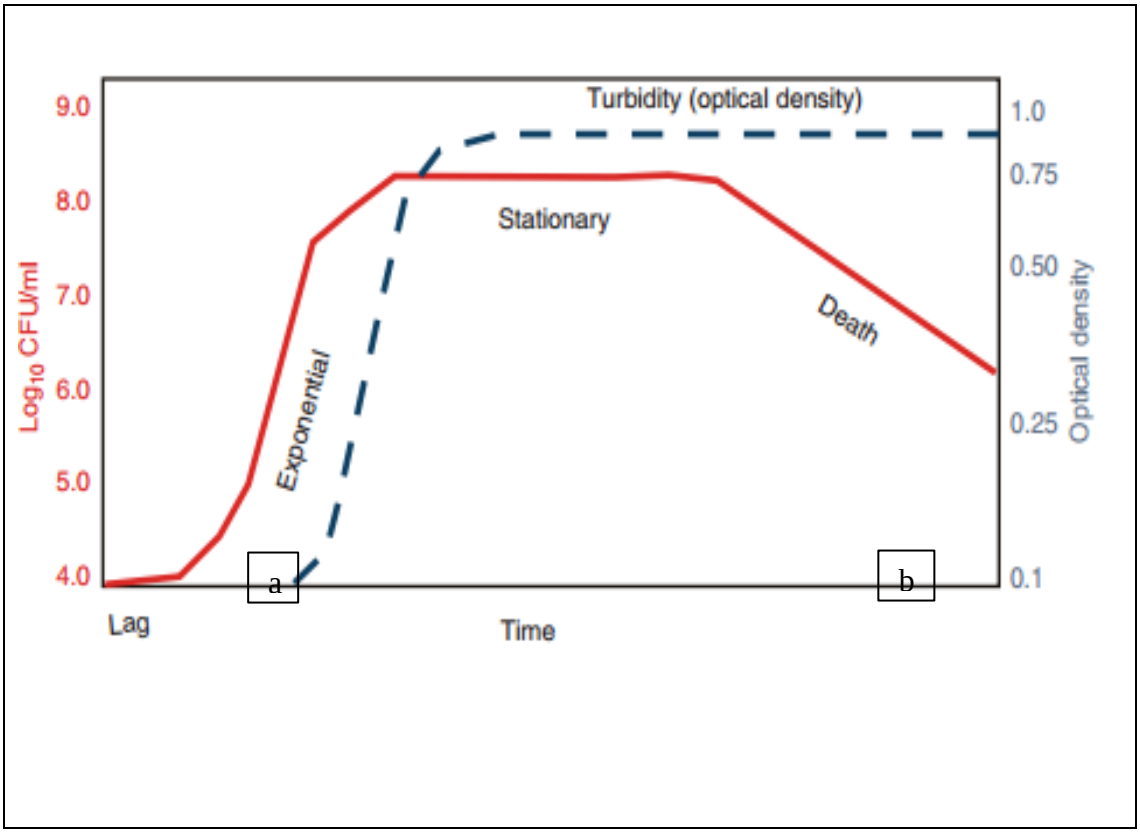


Figure 2-13: Typical growth curve adopted from (Maier, 2009)

The lag phase

Yates and Smotzer (2007) identified that subsequent to the placement of feed in the inoculum, there is no growth (lag phase). This lack of growth is attributed to physiological adaptation of the microorganism to the new environmental conditions (Maier, 2009). The lag phase usually take place from minutes to several hours and is highly dependent on the type and size of feed and inoculum (Maier, 2009). Thus the length of the lag phase could be controlled by ensuring that the inoculum size is adequate compared to the feed, the inoculum is active and variation in the operating conditions such as temperature and mixing is minimised (Maier, 2009).

Exponential phase

Sequel to the initial doubling of the microorganism there is a transit from lag to exponential phase. In this phase, the microorganism undergoes the most rapid growth possible such that the rate of increase of cells in the culture is proportional to the number of cells present at any particular time (Maier, 2009). This exponential growth could be represented with equation 2-10;

$$\frac{dX}{dt} = \mu X \ (Batch\ reactor) \qquad \text{Equation 2-10a}$$

$$\frac{dX}{dt} = \mu X - DX \ (Continous\ reactor) \qquad \text{Equation 2-10b}$$

Rearranging and integrating equation 2-10a, we have;

$$X = X_0 e^{\mu t} \qquad \text{Equation 2-11}$$

Where;

μ = specific growth rate of bacteria (1/time)

X = Final Concentration of microorganism (mg/L)

X_0 = Initial Concentration of microorganism (mg/L)

t = time

D= Dilution rate = 1/time

Stationary phase

In the stationary phase, the net cell growth is zero as any cell growth is balanced by equal number of death cell. In this phase, lack of nutrient result to prevalence of endogenous growth with dying microorganisms lysing to provide source of nutrient. In addition, the build of waste to inhibitory or

toxic level could lead to stationary phase whereby the microorganism could not effectively synthesise the feedstock (Maier, 2009).

Death phase

Most microorganism die compared to those growing. The death phase could be exponential in a batch system; however, it is always slower compared to exponential phase. The death phase is represented as in equation 2-12;

$$\frac{dX}{dt} = -k_d\, X$$ Equation 2-12

Where

k_d = specific death rate (1/time)

2.8.2.2 Simple monod kinetics

Monod kinetics was one of the first and simplest descriptions of entire phase of bacteria growth that was developed empirically through series of experiment. The Monod kinetic described the relationship between specific growth rate and substrate concentration if there is absence of any other form of toxicity or inhibition (Monod, 1949) as shown in equation 2-13.

$$\mu = \frac{\mu_{max}\, S}{K_s + S}$$ Equation 2-13

Where

k_s = half saturation constant (mg/L)

S = substrate concentration (mg/L)

μ_{max} = Maximum specific growth rate (1/time)

The constants μ_{max} and K_s (substrate concentration at which the bacterial growth is half the value of μ_{max}) are intrinsic functional properties of a particular type of microorganism and are also dependent on the type of substrate and temperature of the system.

Incorporating microbial growth in Monod kinetics, we have equation 2-14;

$$\frac{dX}{dt} = \mu X \qquad \text{Equation 2-14}$$

Since substrate utilisation is related to bacterial growth by a constant Y (microbial yield), Monod kinetic could be presented as equation 2-15;

$$\frac{dS}{dt} = \frac{1}{Y}\frac{dX}{dt} \qquad \text{Equation 2-15}$$

Where

Y= microbial yield (mass of microorganism/mass of substrate)

Y varies depending on the substrate type and the intrinsic functionality of the microorganism.

The Monod kinetics is mostly adopted for most mechanistic model, however, there have been various other model that have been developed based on bacterial growth and substrate consumption such as the first order model, Chen & Hashimoto model, etc.

Recently, a simple model for anaerobic digestion of food industry waste was developed based on the assumption that the organic matter (in terms of COD) is composed of a hardly and an easily biodegradable COD fraction, which is consumed through different Monod kinetics towards methane production (Antonopoulou, et al., 2015). The kinetic parameters were estimated by carrying out Biochemical methane potential (BMP) tests and continuous stirred tank experiments. The kinetic model involving only one additional parameter (biomass retention factor) was developed and used to predict the

behaviour of a periodic anaerobic baffled reactor (PABR) fed with the same waste (Antonopoulou, et al., 2015). Despite the structural and operational differences of the two anaerobic digesters, the simple model was able to satisfactorily predict methane production and COD consumption of both bioreactors over a wide range of operating conditions.

However, the challenge with using Monod Kinetics is that it is best suited for steady state thus may not be suitable for dynamic variation that take place in the AD process prior to and during foaming occurrence. In addition, most of the parameters are determined using batch system which functions differently from full scale anaerobic digestion process that is continuous. Furthermore, research has discovered that as AD recovers from system perturbation such as foaming, there are always changes in the microbial mix in the digester. Accommodating this will be a challenge for a model whose parameters have been determined based on a specific microbial mix to assist the AD to recover easily and continue operation efficiently.

2.8.2.3 Hill Model

Hill (1983) identified the need to achieve optimal design of AD via the development of a simplified model that can predict dynamic process both for microbial wash-out and inhibition with few inputs that can be easily determined for complex digester feed. The basic kinetics was based on the Monod kinetics. Conventional Monod kinetics require basic parameters resulting in great number of inputs defining the chemical and physical features of the influent. Hill (1983) realised that Monod kinetic parameters could be lumped according to the different type of waste thereby simplifying it to the extent that inputs for complex substrates could be easily determined by reducing it to an organic base without drastically reducing the potential for the model to accurately predict the system operation.

The organic base is made up of volatile substance and VFA that the acidogenic and methanogenic bacteria act on respectively. For all type of waste, Hill (1983) assumed that the volatile substance and VFA are the same thus the kinetic parameters of the core mathematical relationships will be constantly denoted by biodegradability coefficient (B_0) and acid coefficient

(AF) respectively as enlisted in Table 5. Thus, the determination of B_0 and A_0 were the only two parameters required to characterise the digester feed for the different type of waste, the model is simplified to four equations on mass balance and two equations on microbial growth culminating to a total of only six equations as illustrated in equations 2-16 to 2-21

Biodegradable Volatile Solid (BVS) mass balance equation Equation 2-16

$$\frac{dS}{dt} = \frac{S_0 - S}{\theta} - \frac{\mu_a X_a}{Y_a}$$

Where

S_0 = concentration influent BVS

S = concentration effluent BVS

t = time

θ = detention (d)

X_a = concentration of acid forming bacteria (g organism/ L)

μ_a = specific growth rate of acid forming bacteria (1/d)

Y_a = yield coefficient of acid forming bacteria (g organism/ g BVS).

VFA mass balance; Equation 2-17

$$\frac{dS_a}{dt} = \frac{S_{0-a} - S_a}{\theta} + \frac{\mu_a X_a}{Y_a}(1 - Y_a) - \frac{\mu_c X_c}{Y_c}$$

Where

S_{0-a} = Concentration of influent VFA (g VFA/L).

S_a= Concentration of effluent VFA (g VFA/L).

μ_c = Specific growth rate of methane-forming bacteria (per day).

X_c= Concentration of methane-forming bacteria (g organism/L).

Y_c= Yield coefficient of methane-forming bacteria (g organism/ g VFA).

Acid-forming bacteria mass balance; Equation 2-18

$$\frac{dX_a}{dt} = \left(\frac{\mu_a - K_{da} - 1}{\theta}\right) X_a$$

Where

K_{da} = Specific death rate of acid forming bacteria (per day)

Methane-forming bacteria mass balance; Equation 2-19

$$\frac{dX_c}{dt} = \left(\frac{\mu_c - K_{dc} - 1}{\theta}\right) X_c$$

Where

K_{dc} = Specific death rate of methane forming bacteria (per day)

Monod kinetic growth equation for the acid-forming bacteria

Equation 2-20

$$\mu_a = \mu_m \left(\frac{1}{\frac{K_{sa}}{S} + 1 + \frac{S_a}{K_{ia}}}\right)$$

Where

μ_m = Maximum specific growth rate of acid-formers (per day).

K_{sa} = Monod half-velocity constant (gBVS/L).

K_{ia}= VFA inhibition coefficient for acid-forming bacteria (gVFA/L)

Monod kinetic growth equation for the methane forming

$$\mu_c = \mu_{mc}\left(\frac{1}{\frac{K_{sc}}{S} + 1 + \frac{S_a}{K_{ic}}}\right)$$ Equation 2-21

Where

μ_{mc} = Maximum specific growth rate of methane formers (per day).

K_{sc} = Monod half-velocity constant (gVFA/L).

K_{ic}= VFA inhibition coefficient for acid-forming bacteria (gVFA/L)

Table 2-6: adapted from Hill (1983) on the biodegradability and acid coefficient for Hill simplified dynamic model based on Monod kinetics.

Waste type	$\boldsymbol{B_0} = \frac{\boldsymbol{g\ VS\ destroyed}}{\boldsymbol{g\ VS\ added}}$ $\boldsymbol{as\ \theta \rightarrow \infty}$	$\boldsymbol{A_0} = \frac{\boldsymbol{g\ Violatile\ fatty\ Acid\ (VFA)}}{\boldsymbol{g\ Biodegradable\ Volatile\ Solids\ (B}}$
Swine	0.9	0.07
Beef (confinement)	0.65	0.05
Beef (dirt)	0.56	0.05
Dairy	0.36	0.05
Poultry (Broiler)	0.70	0.20
Poultry (Layer)	0.87	0.20

Based on the six equations, there are eight kinetic parameters and two yield coefficients that must be determined while applying the model. They include; μ_a,μ_c, K_{ia}, K_{ic} ,K_{da},K_{dc},K_{sa}, K_{sc}, Y_a, and Y_c. It is quite pertinent to note that these parameters are true constants for simulation purposes as they do

not change with variation in the waste. Temperature dependent function was applied in the determination of μ_a and μ_c. It was assumed that K_{da}, K_{dc} equals $0.1\mu_m$ while K_{ia}, K_{ic} , K_{sa}, K_{sc}, Y_a, , and Y_c were determined by computer iteration and choosing the best fit of real operating data as 12.0 g VFA/L, 6.0 g VFA/L, 9.0 g BVS/L, 2.0g VFA/L, 0.1 and 0.005 respectively.

Once the constants are determined, four inputs are required for the mathematical model which include; influent total volatile solids (g VS/L) temperature (℃), detention time (θ) and waste type. The model is able to predict volumetric methane productivity, volatile solids reduction, and volatile solids/VFA concentration in the effluent (Hill & Nordstedt, 1980). However, prediction of system perturbation such as foaming was not developed in this model.

2.8.2.4 ADM1

International Water Association (IWA) Task Group for Mathematical Modelling of Anaerobic Digestion Processes established in 1997 was made up of professionals from different anaerobic processing sectors. The aim of the task group was to develop a standard model that forms a common platform for dynamic simulations of various anaerobic processes (IWA Task Group for ADM 1, 2002). Anaerobic digestion model 1 (ADM 1) published in 2002 was meant to promote extensive application of simulation from municipal wastewater and sludge treatment systems to specialised industrial applications and as a useful tool for research, design, operation and optimisation of anaerobic processes worldwide (IWA Task Group for ADM 1, 2002).

ADM 1 structured steps of biochemical and physicochemical processes include the following key stages: disintegration, hydrolysis, acidogenesis, acetogenesis, and methanogenesis. The highlights of the model include (IWA Task Group for ADM 1, 2002):

1. ADM1 involves 19 processes, 24 components and 56 stoichiometric and kinetic parameters (Batstone, et al., 2002)

2. All extra-cellular (hydrolysis) were developed based on first order model while all intra cellular biochemical reactions were assumed to follow Monod substrate uptake kinetics.

3. Gas-liquid transfer and ion association/dissociation were the process assumed to be controlling physicochemical reactions.

4. In ADM 1 model, pH, hydrogen, and ammonia inhibition were considered but not foaming occurrence.

ADM 1 has been successfully tested on variety of systems from full-scale waste sludge digestion to laboratory-scale thermophilic high-rate UASB reactors during and after the model development (Batstone, et al., 2006). The ability to achieve such success was because the structure of the model was developed to encourage specific extensions or modifications with outputs including common process variables such as gas flow and composition, pH, separate organic acids, and ammonium (IWA Task Group for ADM 1, 2002).

Although complex models like ADM1 are very efficient for process simulation, they are significantly limited for process control. The key limitations for ADM1 were centred on over-specificity and omission of some key processes in an attempt to develop a more generic and adaptable model. Besides, in large-scale digesters, it is difficult to relate the assumptions for the model development to field operations such as inability to achieve ideal mixing in large scale digester compared to the assumption of a complete digester mixing in ADM1. Moreover, the difficulty in generating the many input parameters required for the various stoichiometric and kinetic equations has been an issue of great concern.

Detailed analysis and determination of concentrations for soluble and particulate carbohydrates, protein, lipids, and inert components fed to the AD are crucial for using ADM 1. Unfortunately, it is not practical to obtain these results regularly. Although typical values of these parameter have been provided in Batstone et al. (2002) for municipal wastewater, with slight modifications suggested in Batstone et al. (2006), issues still arise if the AD operation ceases to be steady and turns dynamic as applicable during foaming occurrence. Thus, there is need to develop a more robust model that could be

efficiently utilised in the process control of AD operation such that the model could efficiently withstand both steady and dynamic state as well as be able to restore the digester operation in the face of system perturbation very quickly.

Overall, the preference for mechanistic model centres on its ability to predict beyond existing operating data. However, that does not go without some draw back as challenges due to deviation of the existing system from the configuration of the physical system from which the parameters are derived could arise during model calibration (Rabee & Adeloye, 2012). Modelling of foaming in anaerobic digester using mechanistic model has not been achieved as there is non-existence of a clear understanding on foaming occurrence to form the basis of the mechanistic model development (Kanu, et al., 2015). In addition, Saye & Sethian (2013) identified foaming as a unique example of multiscale physics of which the underlying physics takes place over vastly different time and space scales limiting the ability to derive models that quantitatively predict foam evolution.

2.8.3 Empirical model in anaerobic digestion

Empirical models are based on some external features of a system rather than on the mathematical expression of the functional mechanism and structural connectivity of the system. Thus, they are very useful especially in situations where the existing knowledge about the system to be modelled does not suffice to allow the development of a mechanistic model as applicable in the case of anaerobic digester foaming.

Empirical models tend to exhibit a more refined control with highly flexibility and very efficient in local prediction. However, it lacks the ability to be applied to any global prediction as it was developed based on data from the system which may differ from other systems. On the other hand, because empirical models incorporate the existence of uncertainties in the system, they provide better predictions from the practical perspective. The empirical model includes but not limited to artificial intelligence techniques such as expert systems, Fuzzy Logic, artificial neural networks, speech recognition, machine vision, etc.

Since empirical models are data driven, it is faced with the challenge of complex data analysis and making best selection amongst various variables to achieve optimal result. Thus, to enhance the predictive efficacy of the model, data collection process must be properly developed such that collected data is a true representative of the process to be modelled. The collected data must be further processed by removing irrelevant information (outliers) and discovering useful patterns. By using the most relevant variables, the data could be presented in a reduced number of dimensions that will form the inputs for the model. Using historical data, dimensionality reduction can be achieved for a given set of data through features selection and features extraction (Rustum & Adeloye, 2009) which are vital steps in data preparation/transformation. On the other hand, for most experimental design, due to complexity of some process and high cost of data collection, synthetic data generation may become essential in certain circumstance to create sufficient data for modelling.

2.8.3.1 Artificial intelligence

Artificial intelligence (AI) involves the simulation of human intelligence by machines, especially the computer. Consequent to great advances in computational ability has led to huge improvements in the development and application of AI. The wide adoption of AI stems from the ability to efficiently represent complex process with non-linear characteristics without the onerous task of using deterministic non-linear mathematics and equations. Artificial neural network and fuzzy logic system are the most commonly used AI methods (Dalmaua, et al., 2010; Dalmau et al., 2009; Rustum & Adeloye 2012; Arumugam, et al., 2015).

2.8.3.1.1 Artificial Neural Networks

Artificial neural networks (ANNs) are computing algorithms developed based on inspiration from the way and manner the human biological nervous system functions (Haykin, 2009). Interests in ANNs have increased greatly as mechanistic models are not sufficient to model various process due to limited knowledge, ethical, cost and time constraints.

There have been wide application of ANNs to solve complex problems in various fields such as, pattern recognition, identification, classification, speech, vision, acoustics, robotics, image processing, financing, control systems, aerospace, banking, defence, electronics, manufacturing, medical, oil and gas exploration, securities, telecommunications, transportation, real life problems in neuroscience, biological science, earth science, physical science, chemical engineering, civil engineering, structural engineering, translation engineering, etc. (Rustum & Adeloye, 2011).

Specifically, ANNs have been applied successfully to solve water resources problems such as rainfall forecasting (Olsson, et al., 2004), generalised storage-yield- reliability planning (Adeloye & De Munari, 2006), water treatment process (Liu & Kim, 2008), and improved modelling of wastewater treatment primary clarifier using hybrid ANNs (Rustum & Adeloye, 2012).

In AD, ANNs have been applied to model, and subsequently control, methane production using several FFNNs (Holubar, et al., 2003). At steady state condition, four continuously stirred AD were operated at organic loading rates (Br) of about 2 kg m3/d chemical oxygen demand (COD), and disturbed by pulse-like increase of the organic loading rate (carbon source) added to the basic feed (surplus- and primary sludge) to simulate co-fermentation and to increase the COD. Measured parameters were: gas composition, methane production rate, volatile fatty acid concentration, pH, redox potential, volatile suspended solids and COD of feed and effluent. A hierarchical system of neural nets was developed and embedded in a Decision Support System (DSS). A 3-3-1, 9-3-3 and 9-3-2 FFNN simulated the pH, VFA and gas production with a regression coefficient of 0.82, 0.86 and 0.90 respectively. Controlling a lab scale AD using this tool, methane concentrations of about 60% at a rather high gas production rate of between 5 to 5.6 m^3/ d (Holubar, et al., 2003).

Strik, et al., (2005) applied ANN to predict trace compounds (Hydrogen Sulphide and Ammonia) to prevent the fuel cell poisoning while applying biogas from AD to generate significantly higher electrical efficiency. The models were predicting the trace compounds, even under dynamical

conditions. The determination coefficients (R^2) of 0.91 and 0.83 were obtained for hydrogen sulfide and ammonia respectively. Based on the result, several model predictive control tool strategies were with potential to foresee, control, reduce or even avoid the presence of the trace compounds. (Strik, et al., 2005).

Using spiking neural network-based model for anaerobic digestion process, Sciuto, et al., (2016) performed a long-term prediction of the concentration of biogas (CH4 and CO2) at the 100th day of the process, by analysing the concentration evolution of 6 measurable marker-molecules namely CH_4, CH_4S, CO_2, H_2, H_2S and NH_3 during the first 10 days of the process. The model was validated using a small domestic digester which showed an excellent agreement between the predicted values and those obtained with the digester (Sciuto, et al., 2016).

Other research where ANN has been applied to optimise AD include but not limited to the following: modelling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. (Qdais, et al., 2010); modelling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm (Kana, et al., 2012); modelling of pH response in continuous anaerobic acidogenesis by an artificial neural network (J Horiuchi, 2001) and simulation/optimisation/instrumentation of agricultural biogas plants (Wolf, 2013). A comprehensive review of instrumentation and control of anaerobic digestion process have been carried out by Jimenez, et al. (2015).

ANNs are adaptive computer systems whose structure can change due to variation in data flow as the network learn from the data to enable it to understand the system to be modelled. Thus, ANNs can gain knowledge from learning experience that enables it to make useful predictions when a new set of input is presented to it. The knowledge learnt from the data presented to the ANN is stored in the interconnections between the network elements (Arbib, 2003). This process of adapting to changes in data is known as the learning process (Haykin, 2009).

Simple neural network models were developed by McCulloch and Pitts (1943) based on their understanding of neurology making them the founding fathers of neural networks. Subsequently, there have been a lot of development in ANN with excellent description on the history, theory and mathematical basis of ANNs from various literature (Eberhart & Dobbins, 1990; Rustum & Adeloye, 2011; Arbib, 2003; Haykin, 2009; Beale, et al., 2015).

Thus, the focus in this study is on the key components of ANN and how they influence the efficacy of the network. A simple work flow for ANN is as shown in figure 2-14.

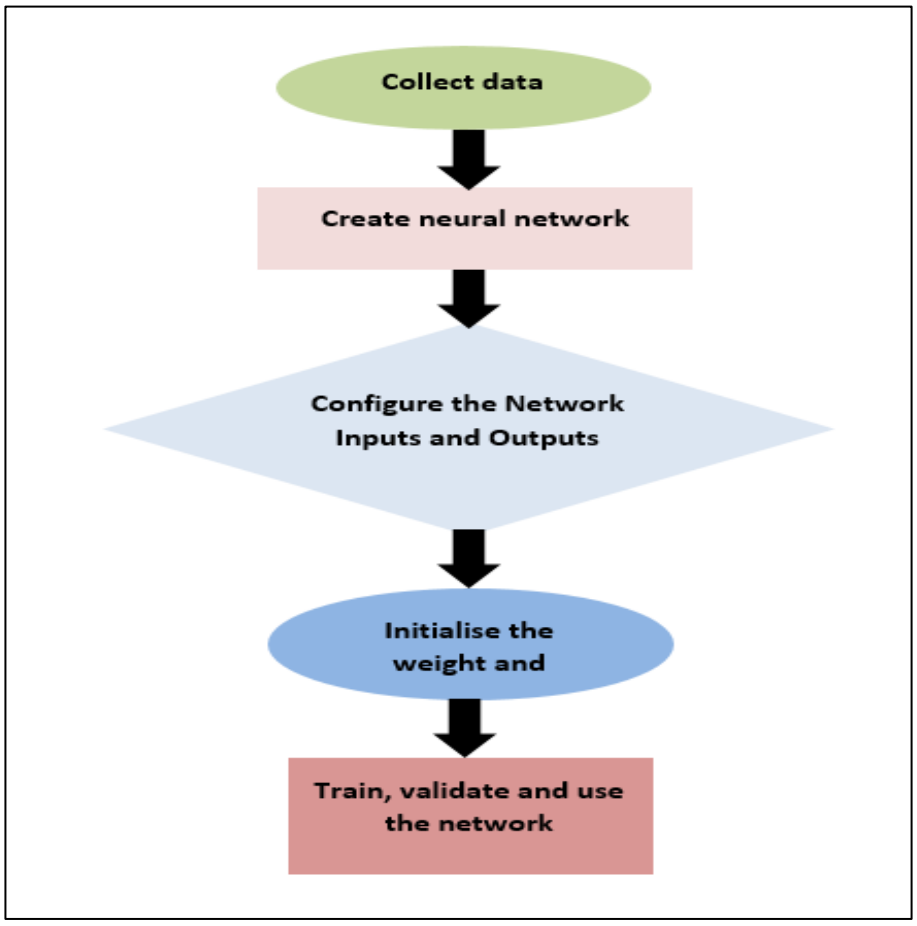


Figure 2-14: Work flow for ANN adapted from (Beale, et al., 2015)

a. Data collection and pre-processing

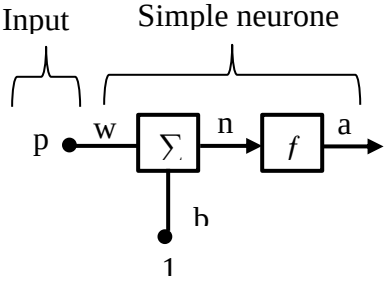
The concept of data collection has been discussed in session 2.8.2.1, however, for more efficient NN training, the collected data is usually pre-processed as expedient depending on the nature of data collected. Data pre-processing

could be in the form of identifying and marking off missing values, normalising the data to achieve a given distribution or dividing the data into sects (training, testing and validating). The division of the data into the training, testing and validation data is essential as the training set is used in computing the gradient and updating the network weight and biases. The training and validation error usually, decreases at initial phase of the training, however, the validation error begin to increase once over-fitting, consequently, the network weights are set are the minimum of the validation set error.

In MATLAB, these pre-processing functions are carried out on the input and output by the network and they become part of the network object allowing new data coming into the network to be pre-processed the same way.

b. Creating a Neural Network

A typical ANN is developed based on the functional operations existing in a simple neuron as shown in figure 2-15. These functional operations are distinct and include; weight function, net input function and transfer function. In summary, for such a simple neuron, the scalar weight w multiplies the scalar input p to form the product wp (scalar), the weighted input wp is added to the scalar bias b (which is much like a weight with a constant input of 1) to form the net input n. The bias b provides an additional variable that can be adjusted to obtain the desired network performance. The net input is then passed through the transfer function f to produce the scalar output a.



$$a = f\,(wp + b)$$

Figure 2-15: A simple neurone

Typically, the transfer function is chosen by the designer while w and b are adjusted by some learning rule so that specific target could be met through the input/output relationship of the neuron.

c. Configure the network and initialise the weight

The simple neuron is applicable to only one input and output variable. However, in real life most analysis involves lot of variables. In such circumstance, the simple neurone could be further developed to handle vector inputs as shown in figure 2-16.

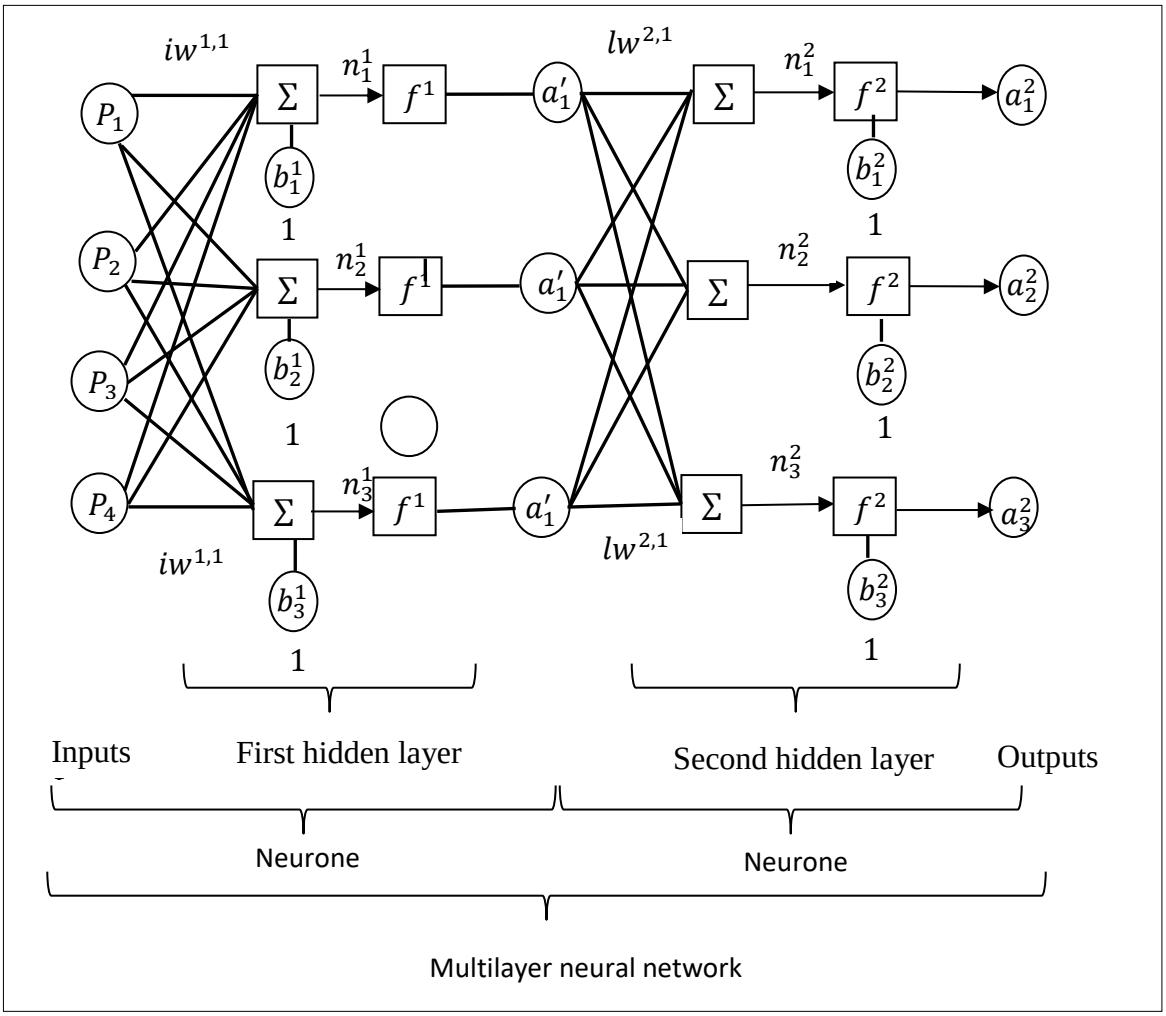


Figure 2-16: Neural network structure

Consequently, two or more neurons can be combined in a layer. A layer includes the combination of weights, the multiplication and summing operation, the bias b and the transfer function f. It should be noted that, the

number of neurons in a layer do not need to equal the number of inputs to that layer. A typical network consists of sequence of layers with a connection weights between successive layers. Usually, these layers are known as input, hidden and output layers. Each hidden or output layer receives values through the connections from previous layer to combine to a single scalar value. The scalar value is passed through a transfer function, also known as activation function.

Both weight and bias are essential component of the neural network as they allow for the network parameters to be adjusted in order to achieve the desired behaviour. There are different weight functions such as; convolution weight function, dot product weight function, negative distance weight function, normalized dot product weight function and scalar product weight function. However, the product of a weight times the input is the most commonly used. The distance functions could also be used as weight functions such as box distance function, Euclidean distance weight function, link distance function and Manhattan distance function.

The transfer function enables the neural network to handle non-linearity as well as limits the permissible amplitude range to finite values of [-1, 1] in the output signal (Haykin, 2009). As shown in figure 2-18, there are various transfer function with the choice of transfer functions to adopt depending on the complexity of the application and expected output. The most widely used transfer functions are: hard limit transfer function, linear transfer function, and sigmoid transfer function (Arbib, 2003).

The hard-limit transfer function limits the output of the neuron to either 0, if the input argument n (wp+b), is < 0, or 1, if n is ≥ 0 as can be seen from equation 2-22

$$f(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$$ Equation 2-22

The linear function is commonly used as it allows the output to be any value as shown in equation 2-23

$$f(n) = n$$ Equation 2-23

The output of the sigmoid function is usually the shape of 'S' and ranges between 0 and 1. The most widely used is the logistic as in equation 2-24

$$f(n) = \frac{1}{1+e^{-n}}$$ Equation 2-24

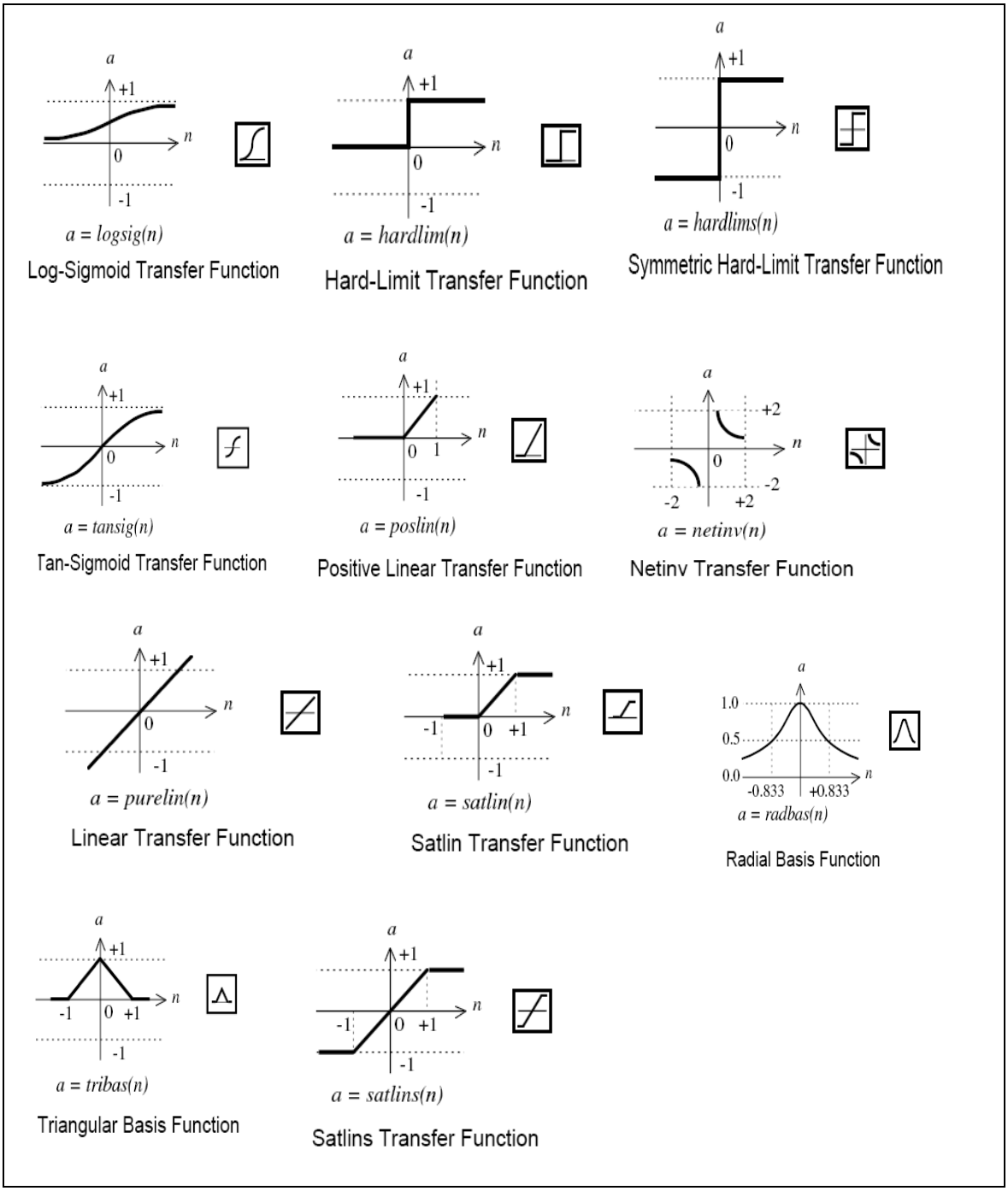


Figure 2-17: Graph of Transfer functions adapted from (Beale, et al., 2015)

The structure of ANN is based on how the inter neuron connections are arranged. Structurally, ANN are grouped into feed-forward neural networks (FFNNs) and recurrent neural networks (RNNs) (Haykin, 2009)

In the FFNNs, information is passed from the input layer to the final output layer in one direction. Therefore, the FFNN is only capable of statically mapping the input vectors to their corresponding targets (Rustum & Adeloye, 2011). Nevertheless, it is widely used due to its ability to model complex functional relationship between the given input and output data sets by learning from examples.

In the RNNs, there is existence of feedback connections within the network either between layers and/or between neurons, a feature that enable RNNs to be dynamic. Consequently, the output at time *t* is dependent on the previous output or state of the neurons within the network as the result of the feed-back. The feedback loops involve the use of branches composed of unit time delay operators. FFNN and RNN can be single layered, multiple layers as well as fully or partially connected. In most instances FFNN with two layers can potentially learn any input-output relationship although the higher the number of layer the faster the network performance (Beale, et al., 2015). Thus, it is best to start with two layers and then move to higher number of layers if the performance is unsatisfactory.

d. Training, validating, and using the network

The training of the neural network is achieved using the learning algorithm. The learning algorithm is the procedure by which the weights of the connections are adjusted to achieve a desired output from the network. Through this process, the synaptic weights of the network are modified in an orderly fashion to attain a desired design objective. (Rustum & Adeloye, 2011). The various make-up of the learning processes is shown in figure 2-18.

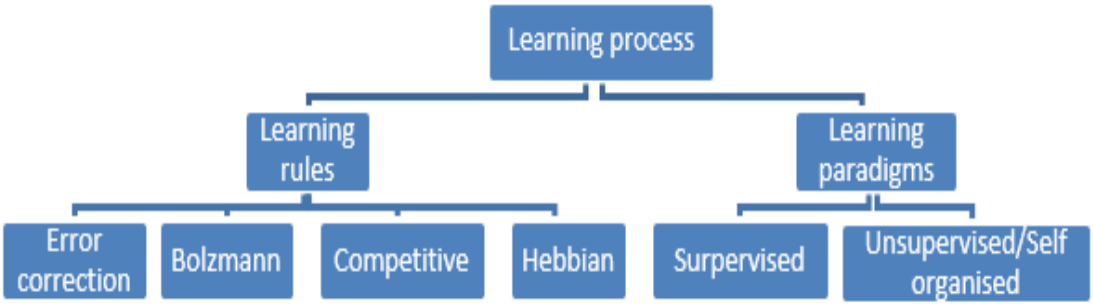


Figure 2-18: Various forms of learning process adapted from (Haykins 2009)

Knowledge representation of the environment for a specified neural network is well-defined by the values taken on by synaptic weights and thresholds of the network which should be adequately optimised (Haykin, 2009). Unfortunately, there is no well-developed theory for evaluating the way in which changes in the network architecture affect the representation of knowledge of the environment.

So, in designing neural network, this can be solved through extensive trials, by comparing different network structures to find the best. Normally, the best mapping is achieved by minimising of the error between network output and the desired output. There exist a lot of algorithms that have been developed over time to proffer solution to different modelling challenges. The fundamental theoretical basis of the different algorithms are extensively discussed in (Haykin, 2009) and so many other literature relating to neural network modelling. Effort is made in this report to highlight the key features of some unique algorithm that are relevant to this study.

2.8.3.1.1.1 Supervised learning algorithms

In supervised learning algorithm, there is adequate knowledge of the environment in the form of input and output variables (target). However, there may be some unknown features of the environment not represented in the input and target variables. Thus, the neural network is trained to learn from the data presented to it based on an error correction learning as shown in figure 2-19. The network parameters are adjusted by combining the effect of the training variables and error signal where the error signal represent the

difference between the expected target and the feedback from the network (Haykin, 2009).

The adjustment of the network is carried out iteratively until the neural network is able to efficiently predict the target such that the knowledge of the environment has been transferred to the neural network (NN) via training and stored in the form of fixed synaptic weights that depict long term memory in human. Since the targets are the optimum action that is necessary to be performed by the network, a measure of network performance is essential and could vary for the different optimisation algorithm.

Summarily, a supervised learning algorithm is capable of reasonably mapping an input-output data if an algorithm designed to minimise the cost function is applied on a set of input-output variables with enough training time (Haykin, 2009). A typical example of learning task involving supervised learning is the function approximation. The different types of supervised algorithm include; perceptron, linear regression model, back propagation etc. (Beale, et al., 2015)

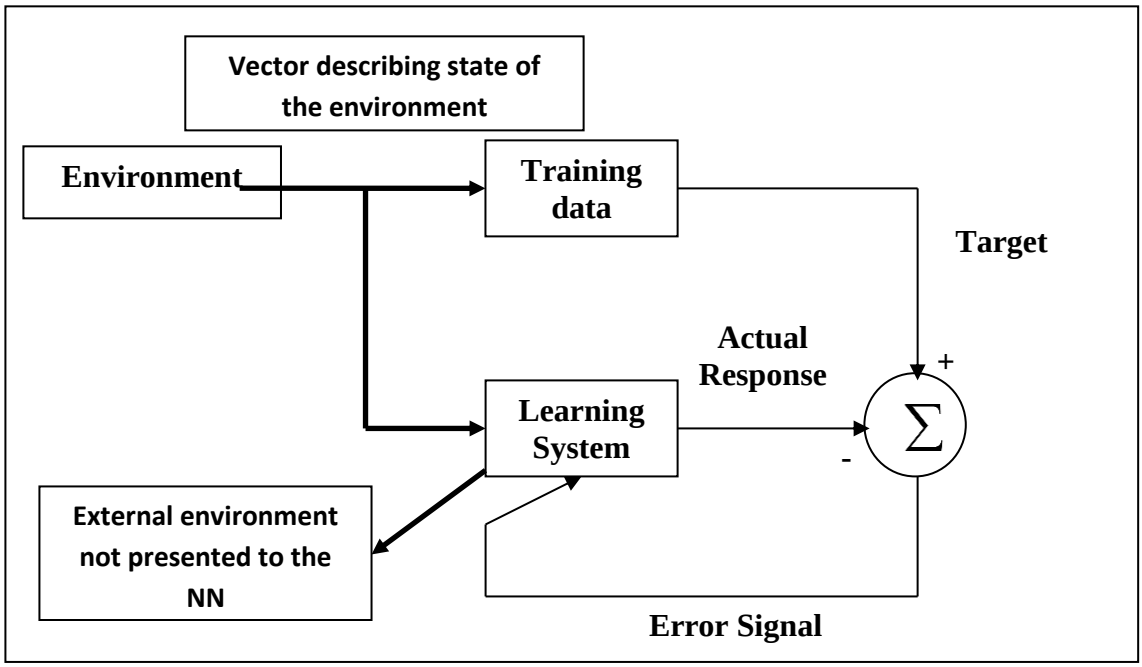


Figure 2-19: Diagrammatic representation of supervised learning (adapted from Haykin, 2009)

Perceptron

Rosenblatt proposed the perceptron as the first model for supervised learning. A perceptron is the simplest form of a NN for classifying linearly separable patterns. It usually made up of a single neurone with adjustable synaptic weights and bias.

The perceptron model is highly dependent on the convergence theorem. Rosenblatt proved the convergence theorem by stating that if the patterns used to train the perceptron are drawn from two linearly separable classes then the perceptron algorithm converges and positions the decision surface in the form of a hyperplane between the two classes. Rosenblatt's perceptron is based on McCulloch-Pitts model of a neuron made up of a linear combiner followed by a hard limited performing the signum function.

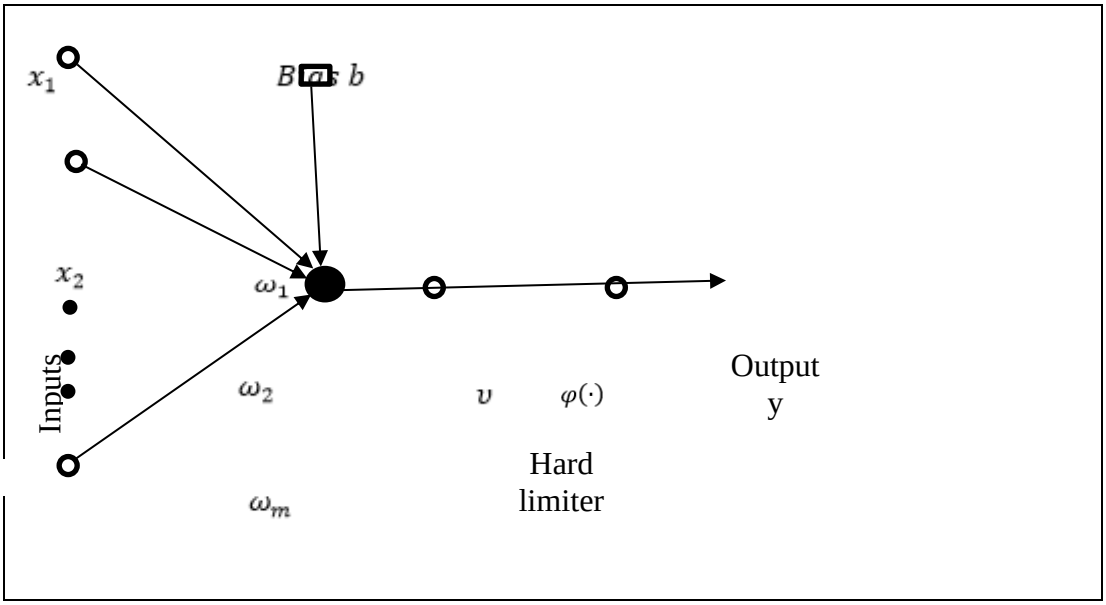


Figure 2-20: Signal flow graph of the Rosenblatt perceptron

As shown in figure 2-20, the synaptic weight is applied to the inputs and summed with the external bias such that the result which is an induced field is applied to a hard limiter producing an output equal to +1 or -1 depending on whether the hard limiter is positive or negative respectively. Thus, the perceptron aims to correctly classify the inputs into two classes of hard limiter as shown in equation 2-25. The synaptic weights are adapted to an

iteration using an error correction rule known as the perceptron convergence algorithm (Haykin, 2009).

$$Hard\ limiter\ (\ell_1, \ell_2) = v = \sum_{i=1}^{m} \omega_i x_i + b \qquad \text{Equation 2-25}$$

Lippman (1987) highlighted the following unique features of Rosenblatt perceptron;

1. The perceptron operates based on the fact that the patterns are linearly separable Thus, when the classes overlap, using perceptron algorithm becomes problematic as the decision boundaries between the different classes may oscillate continuously.

2. The perceptron convergence algorithm is nonparametric as it does not make any assumptions concerning the form of the underlying distributions. It operates by concentrating on errors that occur where the distributions overlap. It may therefore work better if the inputs are generated by non-linear physical mechanisms and when the distribution is heavily skewed and non-Gaussian.

3. The perceptron converge algorithm is both adaptive and simple to implement as its storage requirement is confined to the set of synaptic weight and bias.

Back-propagation algorithm

Training the NN is majorly achieved using the backpropagation algorithm as they are very effective in capturing the non-linear relationships existing between input-output variables in complex systems to generate non-linear regression (function approximation) or pattern recognition.

Although different standard numerical optimisation algorithm could be utilised during training of NN, the gradient of the network performance with respect to the network weight and the Jacobian of the network errors with respect to the weights have shown excellent performance. Backpropagation

computation derived using the chain rule of calculus is used for calculating the gradient and Jacobian of the network (Beale, et al., 2015). Consequently, training of the NN can be either in batch or incremental mode. In batch mode all the inputs in the training set are applied to the network before the weights are updated while in incremental mode, the gradient is computed, and the weights are updated after each input is applied to the network.

Sequel to initialising the network weight and biases, the network training is achieved by tuning the weight and bias to optimise the network performance based on the mean square error (MSE) which is the average squared error between the network output and the target output and calculated as shown in equation 2-26;

$$mse = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \qquad \text{Equation 2-26}$$

Where

t = target output

a = network output

N= number of layers

At each iteration (epoch), MSE is minimised by updating the network weights and biases in the direction in which the performance function decreases most rapidly which is the negative of the gradient. As the training proceeds, the network's weights are adjusted until the network converges. The following training algorithm are used in MATLAB NN tool box; Levernberg-Marguardt, Bayesian Regularisation, BFGS Quasi-Newton, Resilient Backpropagation, Scaled conjugate gradient, Conjugate Gradient with Powell, Fletcher-Powell Conjugate Gradient, Polak-Ribiere Conjugate Gradient, One step Secant, Variable Learning Rate Gradient Descent, Gradient Descent with Momentum and the simple Gradient Descent.

However, Levernberg-Marguardt and BFGS Quasi-Newton were identified to require very low memory and are the fastest training function but less efficient for large networks with thousands of weights performance as they require more memory and computation time. On the contrary, scaled conjugate and resilient back propagation tend to perform better for training large networks while Bayesian regularization training produce better generalization.

On completing the network training and validation, new inputs could be applied to the model to generate a network output. However, some challenges could be met in the performance of backpropagation which includes, inter alia; over-fitting, local minima, and convergence.

Overfitting the model will result in the neural networks been able to provide a very accurate fit of the training data but result in a poor generalisation of unseen (Beale, et al., 2015). Poor generalisation arises when the model is not only modelling the essential dynamic of the system but also unwanted features in the data (noise) (Rustum & Adeloye, 2011).

Using the gradient descent procedure, there is the possibility of the training been trapped in local minima of the network error surface, thereby, exhibiting poor performance during learning which lead to low generalisation capabilities (Rocha, et al., 2007)

Convergence during the learning process using back propagation result to unpredictable and lengthy network such that the length of the network does not necessarily improve the performance of the network (Kamarthi & S., 1999).

Improving results of backpropagation algorithm

Solving network challenge due to overfitting could be achieved by ensuring that large data that is a true representative of the system to be modelled is presented during training provide adequate fit. However, since it is difficult to have a priori knowledge of the size of data, early stopping is normally used. In addition, large sample size decreases the noise effects and improves generalisation of the network.

Early stopping could be applied to all supervised network creation functions, including the backpropagation. This involve dividing available data into three subsets: the training set for computing the gradient and updating the network weights and biases; the validation set for monitoring and detecting data through monitoring the validation error as error on the validation set typically rises at points of over fit; the test data is useful in comparing different models by matching their test error while plotting the test error during the training process could indicate poor division of the data if the test error set reaches a minimum at a significantly different iteration number than the validation set error (Beale, et al., 2015).

2.8.3.1.1.2 Unsupervised learning algorithm

In unsupervised learning, provision is made for a task independent measure of the quality of representation expected from the NN such that the synaptic weights of the network are optimised accordingly (figure 2-21) . Becker (1991) identified that when a network is presented with a specific task independent measure and has become tuned to the statistical regularities of the input data, the NN evolve the propensity to form internal representations for encoding features of the input to create new classes automatically.

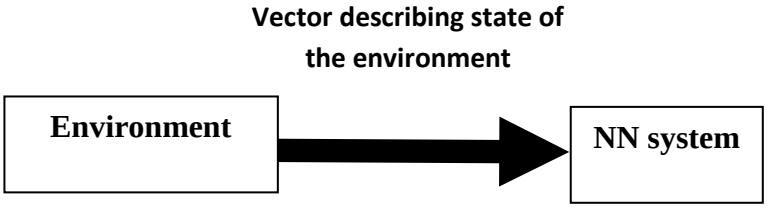


Figure 2-21: Unsupervised learning

Competitive learning rule consisting of input and competitive layers are used in unsupervised learning. The available data is received by the input layer while the competitive layer is made up of neurones that compete with each other based on a learning rule for the opportunity to respond to features contained in the input data. Therefore, the neurone with the greatest total input wins the competition and turns off the other neurones in the network (Haykin, 2009). The resultant effect is that the input data are categorised into

groups or clusters requiring the human contribution and expert to interpret or label the groups appropriately (Kalteh, et al., 2008).Typical examples of unsupervised learning task include pattern association/recognition. Unsupervised learning algorithm that have been widely adopted is the Khonen Self Organising Map (KSOM) (Rustum & Adeloye, 2009; Kalteh, et al., 2008)

Kohonen Self-Organising Map (KSOM)

The KSOM is usually presented as a dimensional grid or map whose units (nodes or neurons) become tuned to different input data patterns. Its algorithms are based on unsupervised competitive learning. During training, only input patterns are presented to the network which automatically adapts the weights of its connections to cluster the input patterns into groups with similar features (Astel, et al., 2007; Kalteh, et al., 2008).

The principal goal of the KSOM is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map. It involves clustering the input patterns in such a way that similar patterns are represented by the same output neurons, or by one of its neighbours (Back et al., 1998). In this way, the KSOM can be viewed as a tool for reducing the amount of data by clustering, thus converting complex, nonlinear statistical relationship between high dimensional data into simple relationship on low dimensional display (Kohonen et al., 1996; Zhang, 2009).

This mapping roughly preserves the most important topological and metric relationship of the original data elements, implying that the KSOM translates the statistical dependences between the data into geometric relationships, whilst maintaining the most important topological and metric information contained in the original data. Hence, not much information is lost during the mapping. Hence, similarities relationship within the data and clusters can be visualised in a way that enables the user to explore and interpret the complex relationship within the data set.

2.8.3.1.2 Strength and challenge of neural network models

Overall, NN have the great potential to be applied in solving complex problem exhibiting strong non-linearity that make it challenging to understand the

system kinetics to develop a mechanistic model. In addition, NN are inherently simple to understand and are developed by using enough actual historical observations from past experience. Thus, NN learn by examples, thereby evolving effective solutions more quickly than could be achieved using traditional mechanistic models, which are entirely reliant on expert knowledge in a field.

Consequently, adopting careful design, NN can generalise on data by producing accurate response to data not presented to the network during training. Nevertheless, the limitation of NN is that they are not good extrapolators as they are unable to simulate outputs outside the range of those they were trained with.

Finally, the choice of the number of layers and the number of hidden neurons in the hidden layers is a major concern in developing NN. Thus, selecting the number of hidden neurones and layers do not follow any specified rule as they are determined through trial and error. In addition, the number of inputs is often unknown and different models with different inputs can be trained in order to select the optimal model. Consequently, the decision of the final model is usually determined by evaluating the trained model using several evaluation criteria on new data and selecting the model that is best suitable for the data available.

2.8.3.2 Fuzzy Logic modelling

Fuzzy logic modelling (FLM) was first applied in the control of a small steam engine by Mamdani and Assilian (Mamdani, 1977; Mamdani & Assilian, 1975). It has also found early adoption as a sludge controller for wastewater treatment plant (Tong, et al., 1980). FLM could be applied to different circumstance such as fuzzy-rule-based systems, fuzzy expert systems, fuzzy associative memory, or fuzzy controllers (Ross, 2004).

Mamdani introduced Fuzzy logic based on Zadeh's theory of fuzzy set (Mamdani and Assilian, 1975). Usually, fuzzy set represent a set that is devoid of a crisp, clearly defined boundary with the elemental components

having only a partial degree of membership. This flexibility highlights the intelligent reasoning of human.

Thus, as shown in figure 2-22, fuzzy theory applies membership degree to any set of objects with values assigned between zero and one rather than been restricted to the integers [0 1] as obtainable in classical sets. These values between 0 and 1, denotes the degree of membership, also called membership value, and varies depending on the membership function (Triangular, Trapezoidal, Gaussian, and Bell-shaped functions) as shown in figure 2-23. The shape of the membership functions depends on parameters, and changing these parameters change the shape of the membership functions.

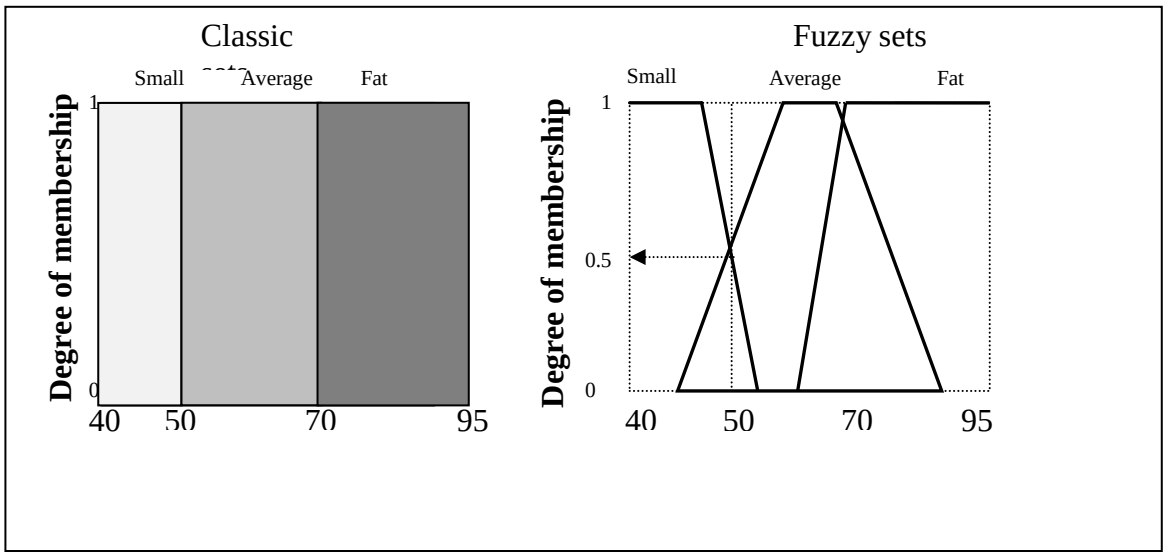


Figure 2-22: Diagrammatical representation of the difference between a classic and fuzzy set

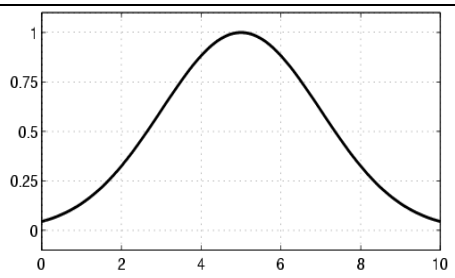
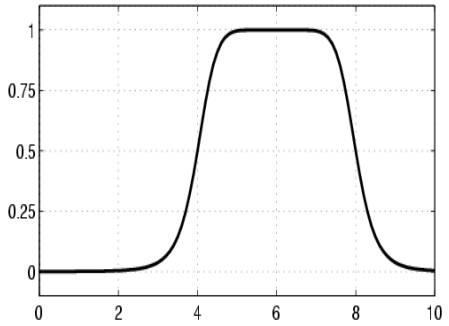
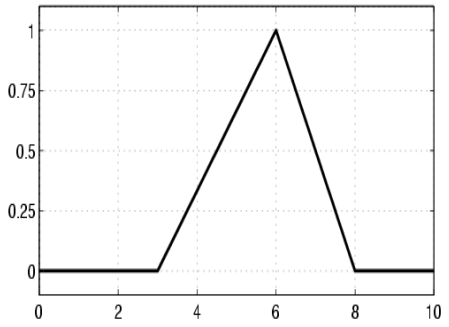
Using the same example in figure 2-22, fuzzy sets F_{small}, $F_{average}$, and F_{fat} of X can be represented as in Equations 2-27, 2-28, 2-29 respectively.

$$F_{small} = \{(x, \mu F_{small}(x)), x \in X \quad \text{Equation 2-27}$$

$$F_{aver} = \{(x, \mu F_{aver}(x)), x \in X \quad \text{Equation 2-28}$$

$$F_{Fat} = \{(x, \mu F_{Fat}(x)), x \in X \quad \text{Equation 2-29}$$

Note: Membership function (MF) depict the degree to which X *is an element of set F*

Name	Description	Example
Gaussian curve built-in membership function	The symmetric Gaussian function depends on two parameters b and a as given by: $f(x;a,b) = e^{\frac{-(x-a)^2}{2b^2}}$	 *gaussmf=f(x,5,2)*
Generalised bell-shaped built-in membership function	The generalised membership function depends on three parameters, a,b, and c as given bye: $f(x;a,b,c) = \frac{1}{1+\left\lvert\frac{x-c}{a}\right\rvert^{2b}}$ where the parameter b is usually positive. The parameter c locates the centre of the curve	 *gbellmf=f(x;2,4,6)*
Triangular-shaped built-in membership function	The triangular curve is a function of three scalar parameters a, b, and c, as given by: $f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$ The parameters *a* and *c* locate the "feet" of the triangle and the parameter c locates the peak.	 *trimf=f(x;3,6,8)*

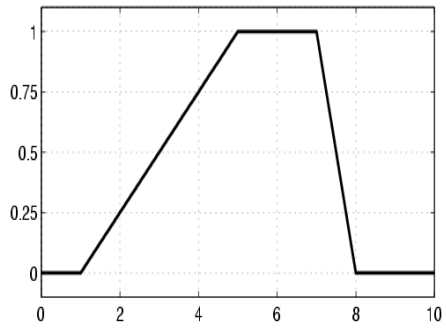
Trapezoidal-shaped built-in membership function	The trapezoidal curve is a function of a vector, x, and depends on four scalar parameters a, b, c, and d, as given by $$f(x;a,b,c,d)=\begin{cases}0, & x\le a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & b\le x\le c\\ \frac{d-x}{d-c}, & c\le x\le d\\ 0, & d\le x\end{cases}$$ The parameters a and d locate the "feet" of the trapezoid and the parameters b and c locate the "shoulders."	 *trapmf=f(x;1,5,7,8)*

Figure 2-23: Fuzzy logic membership functions adapted from (Rustum & Adeloye, 2011)

Fuzzy logic model represents the human ability to oppose, object and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, conflicting information, partiality of truth and partiality of possibility while performing a wide variety of physical and mental tasks in the absence of measurements and computations (Zadeh, 1999). In such framework, solutions are proffered for problems whose source of imprecision is dependent on lack of sharply defined membership class rather than the presence of random variables (Zadeh, 1965).

Consequently, these features offer the possibility of modelling non-linear functions by explaining the reasoning linguistically rather than with numerical quantities in the form of fuzzy rules (Nguyen & Walker, 2006).

Structure of Fuzzy Logic System

Fuzzy logic model (FLM) is a nonlinear mapping from the input to output space based on three concepts namely; fuzzy sets, linguistic variables, and fuzzy if-then rules. The process is carried out as shown in figure 2-24.

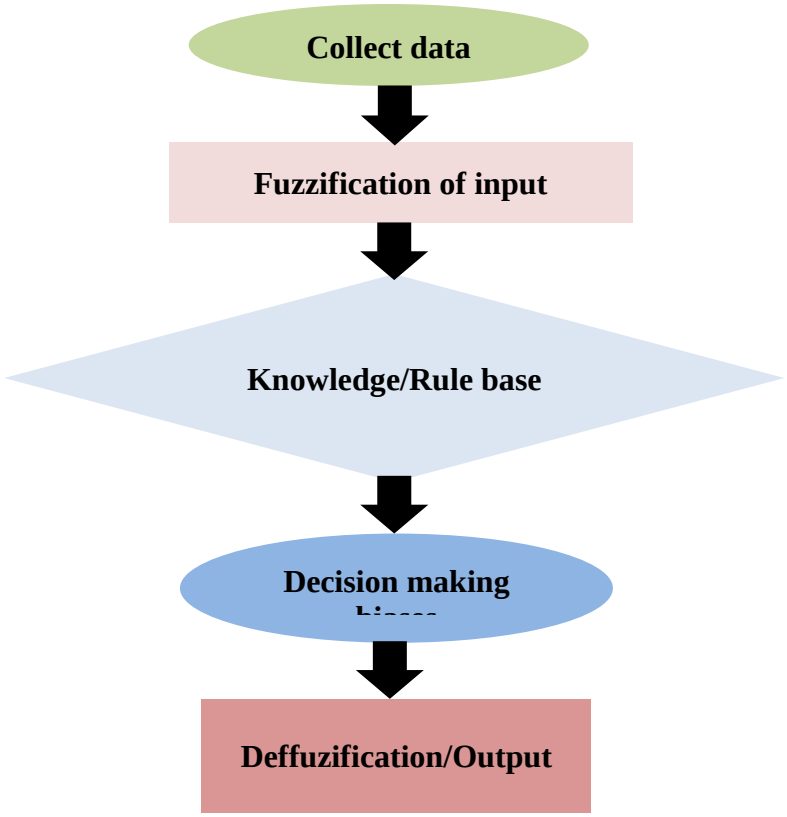


Figure 2-24: Fuzzy logic process

If-then rules

If-then rules are statements used to express conditional statements governing the variables used in FLM. The ability to select the most representative variables will impact the number of rules and the performance of the FLM. Variable selection is highly dependent on expert knowledge as well as relationship existing between the input(s) and desired output(s). Connections between variables are mostly determined using correlation analysis while in some cases other AI classification model such as KSOM could be utilised.

Once the input and output variables are determined, the fuzzy rules based are generated based on expert knowledge and control engineering (neural network fuzzy inference systems). Execution of if-then rules involve the following steps; Fuzzification of input, application of fuzzy operators and implication method.

Fuzzification of input

The input to a FLM is a crisp numerical value from the collected data which is restricted to the universe of discourse of the input variables. As shown in fig 2-222 the universe of discourse for the input variable (weight of people) is 40 – 95 kg. During fuzzification, the inputs are used to determine the degree to which they fit the appropriate fuzzy sets (small, average, fat) via membership functions. Thus, the output of the fuzzification process is a fuzzy degree of membership between the intervals of 0 and1 in the qualifying linguistic set (fuzzy set). The fuzzification process is developed through function evaluation. Triangular membership function was utilised in figure 2-22 with input of 50 kg been assigned a membership degree of 0.5.

Application of operators

The value obtained after input fuzzification is known as antecedent. It is that number that shows the degree to which the fuzzy rule is satisfied by that variable. However, a given rule may have more than one antecedent such that it is necessary to apply the fuzzy operator to obtain a single value for the rule. Assuming we are predicting the body mass index (BMI) of certain group of people with inputs as height (short, normal, tall) and weight (small, average, fat) and the output as BMI (low, high). One of the if-then rules could be;

If height is short and weight is average, then BMI is high

In this rule, there will be two antecedents from height is short and weight is average. Thus, the fuzzy operator will need to join the antecedents to generate a single value of antecedent for that rule. Different operators are used in fuzzy logic but the most frequently used are AND representing the fuzzy intersection/conjunction (minimum and product) OR representing fuzzy union/disjunction (maximum and probalistic) and Not representing fuzzy complement (additive compliment) as shown in figure 2-26;

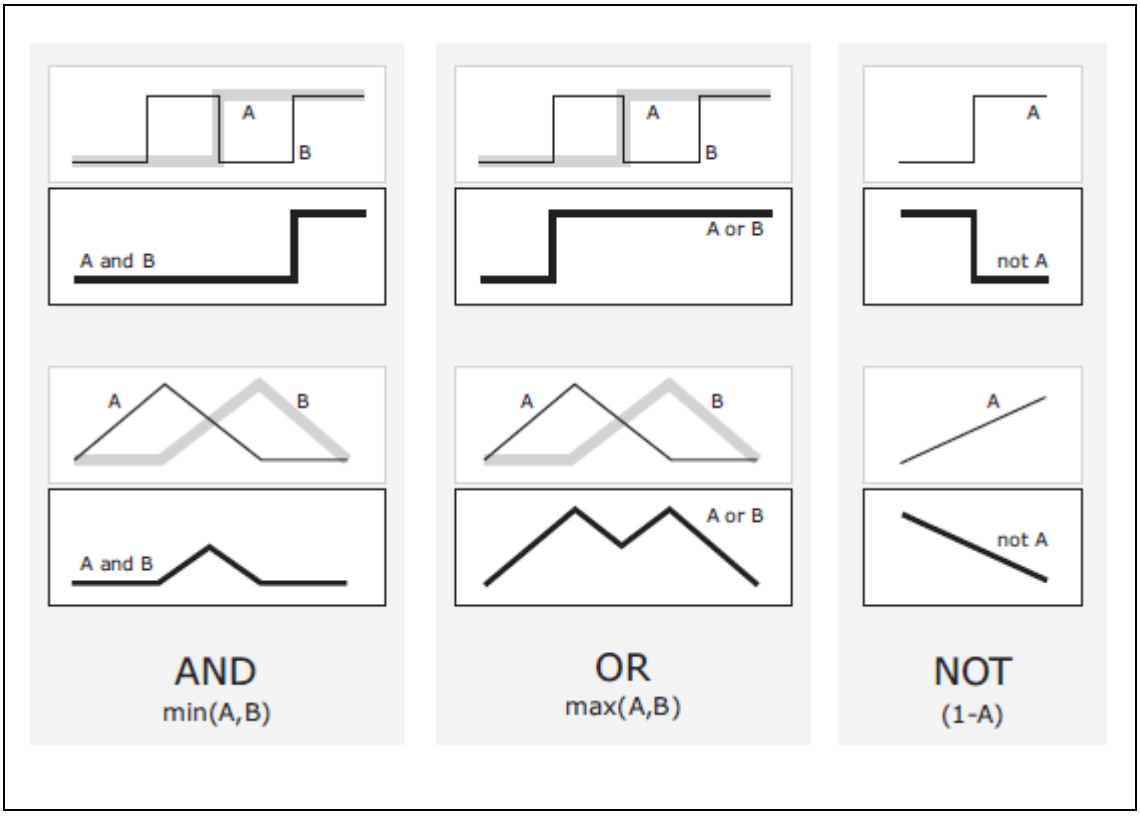


Figure 2-25: Fuzzy logic operators adapted from (Mathworks, 2015)

Consequent to the application of the fuzzy operators from two or more membership values from fuzzified input variables, the output is a single value.

Application of implication

Prior to applying the implication method for each rule, the rule's weight (a number between 0 and 1) is defined and applied to the number generated as the antecedent. In most cases, this weight is 1 thus do not affect the implication; however, there may be instances that a rule may be weighed relative to the other by varying its weight. The result of applying the implication is the generation of a consequent which is a fuzzy set represented by a membership function, which weights appropriately the linguistic characteristics that are attributed to it. Using a function that is associated with the antecedent, the consequent is reshaped by applying AND method (minimum) to truncate the output fuzzy set, and prod method (product) which scales the output fuzzy set.

Aggregating outputs

Aggregation is very useful in combining the output fuzzy sets of the different rules following the implication process to produce a single fuzzy set. Aggregation occurs only once for each output variable resulting to one fuzzy set for each output variable. Aggregation methods include; max (maximum), probor (probabilistic OR) and sum (adding each rule's output set)

Deffuzification

Although fuzziness assists in the rule evaluation during the intermediate steps, the final anticipated output for each variable is a single number. Thus, the aggregate of a fuzzy set comprising a range of output values must be deffuzzified in order to resolve a single output value from the set.

Deffuzification methods include; centroid, bisector, middle of maximum (the average of the maximum value of the output set), largest of maximum, and smallest of maximum (Mathworks, 2015). The most common means of defuzzification is called the centre of gravity method in which the centre of gravity of the fuzzy set is measured and projected to the z-axis to get the crisp result as illustrated by Figure 2-26. The output of this defuzzifier is a number z given by Equation 2-30.

Equation 2-30

$$z = \frac{\int z_i \mu_{zi}(z_i) dz}{\int \mu_{zi}(z_i) dz} \qquad \text{Equation 2-30}$$

where

z = the crisp output,

μ_{zi} =the fuzzy membership value at zi

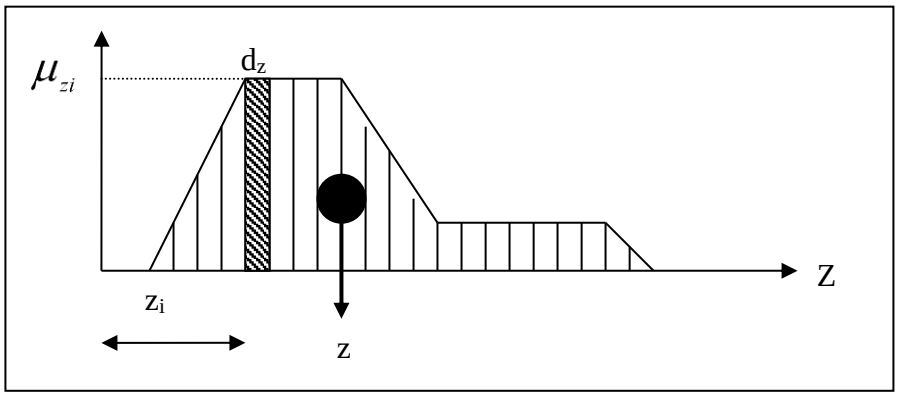


Figure 2-26: Deffuzification using centre of gravity methos

Fuzzy logic models differ mainly due to variation in the if-then rule. They include, Momani, Takagi-Sugeno-Kang's (TSK) fuzzy rules (Sugeno & Kang, 1988; Takagi & Sugeno, 1985) and Neural Network Fuzzy inference system (NNFIS) (Mathworks, 2015).

2.8.3.2.1 Mamdani fuzzy logic model

Mamdani fuzzy logic model (MFLM) applies fuzzy system in the antecedent and consequent of the if-then rule. The selection of the membership function for the variables used in MFLM is dependent on expert knowledge thereby making it difficult to achieve consistency when different experts are involved. Furthermore, as the number of input variable increases, the fuzzy rules increases geometrically requiring sufficient a priori knowledge about the system as well as trial and error process to develop the model. Normally, Mamdani's method has seen more widespread use because it is easy to be understood by human experts.

2.8.3.2.2 Takagi-Sugeno fuzzy model

The Takagi-Sugeno fuzzy model is a specific case of Mamdani fuzzy system as the antecedent of the fuzzy rules are defined with linguistic terms but the consequence is described by a non-fuzzy equation of the input variables which are linear combination of the input variables. The advantage of Takagi-Sugeno's method is that it has better computational efficiency that is very

useful for modelling non-linear systems (Jang & Gulley, 1997; Jang & Sun, 1995).

2.8.2.4.2.3 Neural Network fuzzy inference system

Neural network fuzzy inference system (NNFIS) is a special form of FLM as it provides a method for tuning the parameters based on training data set of input and outputs data using NN training paradigms. The method incorporates the advantages from FLM (data driven) and NN (expert knowledge). Jang (1993) found that combining the supervised learning potential of NN and the heuristic reasoning capability of FLM, NNFIS model is able to learn complex functional relations and generate fuzzy rules at the same time.

Multiple examples of NNFIS were carried out by Jang to validate the model and the results were reported to be comparable with backpropagation NN. However, due to the use of noiseless data sets generated by functional equations, the application of NNFIS on noisy field data was not proven by Jang (Jang et al., 1997). Thus, NNFIS was applied by Miller (2006) to predict the rainfall (precipitation) from noisy weather data (temperature and humidity) and it was found that the model predicted negative values of rainfall on some instance. Miller (2006) suggested that pre-processing of the data, e.g. by replacing the missing values and omitting the outliers may improve the performance of NNFIS.

NNFIS is a Sugeno-type FLM (Sugeno & Kang,1988; Takagi & Sugeno, 1985). NNFIS comprises of a multi-layer feedforward network in which each node performs a particular function on incoming signals (Jang et al, 1997; Chang and Change, 2006). During the development of NNFIS, the membership functions are optimized with the ability to construct fuzzy IF-THEN rules that represent the optimized membership functions. The five layers in the NNFIS are: the input layer, the fuzzification layer, the rules layer, the standardization layer and the output layer. The role of these layers is related their role of the different layers as already discussed in FLM.

Using a given input/output data set, NNFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted)

using either a back-propagation algorithm alone or in combination with a least squares method. Modification of the membership function parameters are carried out according to a chosen error criterion such that NNFIS model could emulate the training data presented to it while adhering to the network rules. If the number of rules is not restricted, a Sugeno model will be able to map any non-linear function (Jang, 1993). On the other hand, the performance of NNFIS is highly dependent on the ability of the training data to be fully representative of the features of the data that it is intended to model.

However, in some circumstances, there may be noise in the collected data limiting the ability of the training data to be a representative of all the features of the data to be presented to the model. In such situations, model validation is helpful such that the NNFIS model is developed using training, testing and checking data collected based on observations of the target system. Model validation is the process by which the input vectors from input/output data sets on which the NNFIS was not trained, are presented to it in order to evaluate the ability for the NNFIS model to calculate the corresponding data set output values. The challenge with the validation is the ability of selecting a data set that represent the target data as well as the trained while been sufficiently different from the training data such that the validation process is not rendered trivial. Thus, there is need for large amount of data to be collected to make it easier to select the testing and checking data set.

During validation of NNFIS, the testing data set check the generalization capability while the checking data set is essential to avoid data overfitting. This is crucial as the model error for the checking data set tends to decrease as the training takes place up to the point that overfitting begins, and then the model error for the checking data suddenly increases. This occurs when testing the trained NNFIS against the checking data and choosing the parameters of the membership function as those associated with the minimum checking error if these errors indicate model overfitting.

2.8.3.2.3 Strength and challenges of Fuzzy logic models

FLM are conceptually easy to understand, flexible and tolerant of imprecise data. As FLM combine knowledge of the system and human expert in developing the fuzzy rule while tuning the parameters as a black-box. This feature offers an advantage over NN models that are solely dependent on the ability of the system to learn from the knowledge presented to it (Sadiq et al., 2004).

The major challenge faced in FLM is the level of expert knowledge for building the fuzzy rules and tuning the parameters in the membership functions which is crucial to achieving the desired performance. In the early applications of fuzzy logic models, the generation of the fuzzy rules and the adjustment of its membership functions were performed manually by trial and error and the best combination found by simulation test (Ross, 2004). Subsequently, NNFIS was evolved to formalise a systematic approach to generate fuzzy rules from an input-output data set based on Sugeno FLM. However, this does not come with its own challenge of requiring noiseless and huge amount of data. Ultimately, the expert knowledge, circumstance of the system to be modelled in terms of complexity and data availability impacts the selection of the form of FLM that will be adopted.

2.9 Conclusion

In this chapter a brief outline of wastewater treatment process was given, which was useful in establishing the process via which sludge fed to the AD is generated, the characteristics of the sludge and how these affect the AD process. A review of the background of AD highlighted the biochemical processes at play in AD and the factors affecting them. Furthermore, foaming and the impact of foaming in AD process were discussed, highlighting the possible causes of foaming in AD process.

'Omics' analysis methods were studied and their importance and relevance as a useful high throughput tool that is vital for developing/validating comprehensive knowledge of the microbial and metabolic constituent of AD content was outlined.

A review of mechanistic and empirical modelling of AD process was presented and discussed highlighting the challenges and strength of each model and the implication for AD modelling. A brief outline on research approach was essential to understanding the research method to be selected that will offer the best benefit to this study.

A major revelation from this chapter is that the AD is a complex biochemical process in which microorganism and operating conditions in the AD play vital role in the efficiency of the process. System perturbation such as foaming occurrence was identified as a challenge militating against the wider adoption of AD process. Thus, there is need to proactively mitigate the occurrence of such system perturbation by developing predictive models that could accurately ascertain the propensity for foaming and form the basis of a potential automated control mechanism.

The most extensive AD mechanistic model (ADMI) did not model foaming occurrence as it is difficult to establish the kinetics of foam occurrence. In addition, the ADM1 requires the determination of about 29 parameters, coefficients, and variables, most of which are not routinely measured in full scale AD, making its adoption as a control process for systems with high variation in the influent and effluent characteristics challenging. Thus, the need to apply empirical models which are data driven and a better representative of the variation in influent and effluent characteristics of the AD to be modelled becomes imperative.

Unfortunately, the absence of any detailed record of the volume of foam by AD operators due to lack of measuring system was identified as a challenge as well as proving the intricacy of the foaming occurrence. This may have limited the inclusion of risk of foaming in AD modelling as it was identified that there was only one AD model that considered the risk foaming occurrence.

Consequently, the need to acquire data through properly designed laboratory experiment, applying synthetic data generation on collected data from laboratory experiment to extend the robustness of the data and increasing

expert knowledge from Omics analysis was identified as crucial for using data driven model to predict foaming occurrence in AD process.

Chapter 3 : Methodology

3.0 Methods

Research methodology is the “overall approach to a problem which could be put into practice in a research process, from the theoretical underpinning to the collection and analysis of data” (Remenyi et al. 2003). It provides answers to research questions by making available means to achieving stipulated goals through which underlying facts and assumptions are captured to justify the rationale behind the research

In developing the research methodology for this study, it was essential to create a conceptual framework that holistically reflect the research design. The conceptual framework as shown in figure 3.1 is designed to show the interaction and interplay of the main objectives in this research and how they are effectively linked to achieving the aim of this study. In addition, it also served as a guide to monitor the research progress during the period of the study. The rest of this chapter contain a detailed discussion of the different stages in the research as shown in the conceptual framework.

3.1 Identify research gap

Extensive review of literature on anaerobic digester operation and existing model were carried out to ascertain factors relating to anaerobic digester foaming and establish knowledge gaps. The understanding gained from the literature review was beneficial to developing the most relevant and up to date approach that was crucial to proffering solution to close up the identified gaps. The literature review was achieved by searching the library/associated libraries and using internet facilities to identify and access the latest publications that are pertinent to anaerobic digester operation, foaming occurrence and modelling.

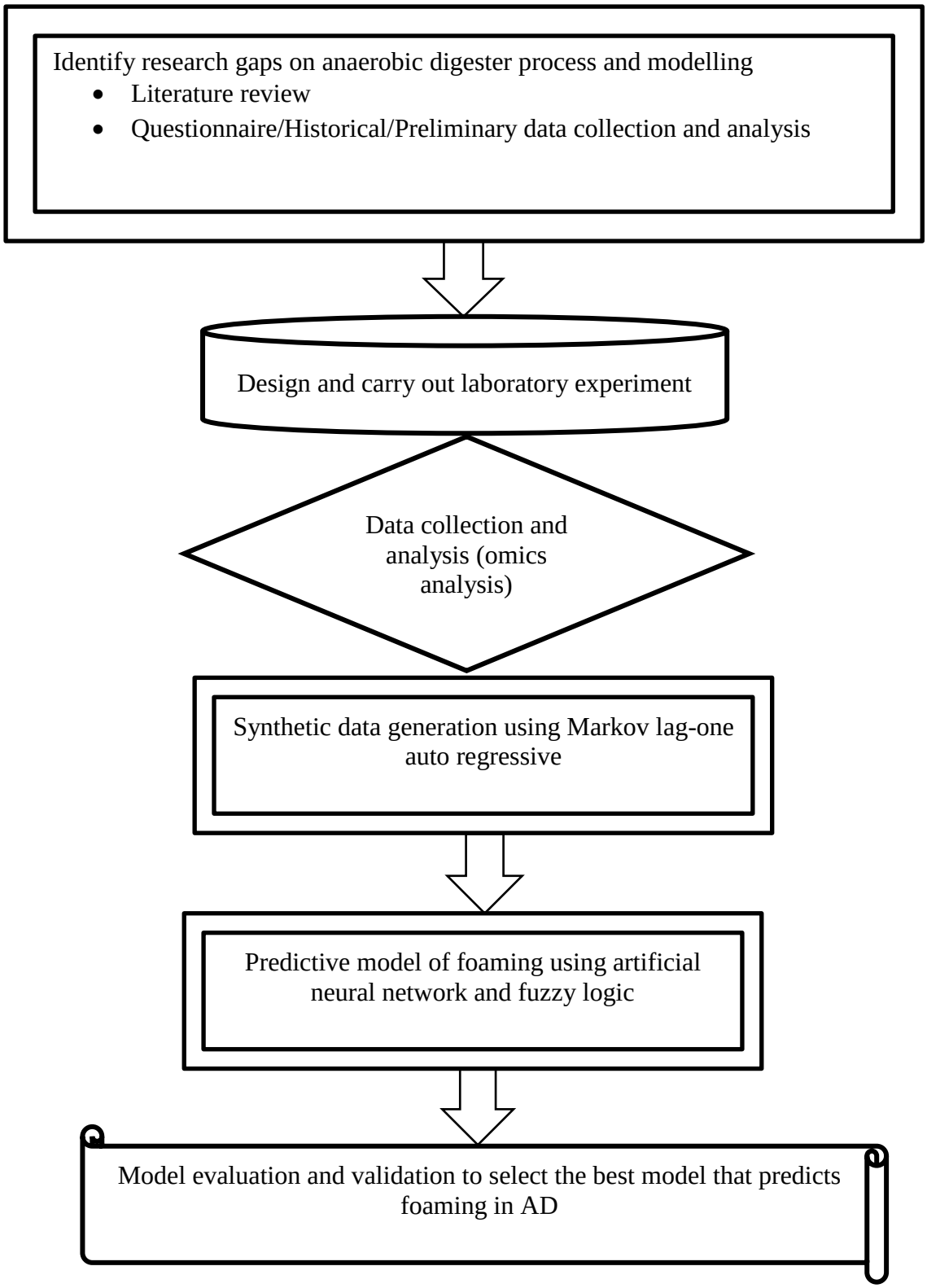


Figure 3-1: Research conceptual framework

3.2 Data Collection

3.2.1 Historical data collection

A questionnaire (appendix 1) was prepared following knowledge gained from the literature review and the gaps identified and sent to all full-scale anaerobic digesters treating sewage sludge in Scotland as shown in table 3.1.

Table 3-1: Anaerobic digesters treating sewage sludge in Scotland

No	Site Name	Treatment type	Managed by
1	Cumnock, East Ayrshire	AD +dewatering	Non PFI (Scottish waters)
2	Stirling	AD +dewatering	Non PFI (Scottish waters)
3	Galashiels & Selkirk	AD +dewatering	Non PFI (Scottish waters)
4	Dalderse (+ Bo'ness), Falkirk	AD +dewatering	Non PFI (Scottish waters)
5	Allanfearn (Inverness)	AD+Dewatering	Veolia Water
6	Hatton (Dundee)	AD+dewatering+drying	Veolia Water
7	Nigg (Aberdeen)	AAD+dewatering	Kelda water services
8	Seafield (Edinburgh)	AD+dewatering	Veolia Water
9	Newbridge (Edinburgh)	AD+dewatering	Veolia Water

This was aimed at collecting data that will offer a deeper insight on the AD process. The questionnaire was sent out through email and feedback on the survey plus historical data was received only from four of the WWTW.

The historical data show that part of the daily routine of operators in monitoring operational efficiency of AD involve keeping record of relevant operating variables such as flow(m^3/d) and % of total dry solids (TDS) in PS from the gravity thickener, flow(m^3/d) and % of TDS in WAS from the centrifugal belt, flow(m^3/d) and % of TDS in mixed sludge (mixture of PS, WAS and in some cases IS) fed to the digester, Temperature, % of methane in the biogas, volume of biogas produced(Nm^3/d), % volatile solid (VS) in the feed and digestate, digester pH, alkalinity ($mg/L\ CaCO_3$) and volatile fatty acid (mg/L VFA) in the digestate.

The data collected also include calculations of OLR and VSR by the operators using equations 2-1 and 2-4 respectively. It is worth mentioning that there were greater consistencies in data for TDS compared to alkalinity, VFA, pH and percentage of methane contained in the biogas. This is because there is no online system for collecting these data which is often onerous to determine regularly at the laboratory scale. The results were discussed in detail in chapter 4.

3.2.2 Preliminary laboratory experiment

The deduction from the analysis of collected data formed the basis for the design of an initial experiment targeted at selecting the most relevant variables causing AD foaming. Consequently, pre-digester (mixed sludge from the holding tank) and digestate sludge samples from Newbridge-Non-foaming (NBP/ NBD) and Seafield-Foaming (SFP/SFD) wastewater treatment plant were collected on weekly basis as the operating conditions permits and analysed for the various factors such as pH, Alkalinity, total surfactants, foaming potential (FP), filamentous index (FI_gram stain), total Volatile fatty acid (tVFA), soluble protein (SP), total solid (TS) and volatile solid (VS).

The data collected were further analysed statistically using the method of comparative design. This involved using the t-test to compare the result of the foaming and non-foaming digesters for the different variables analysed to decipher which of the variables were statistically significant. Thus, the two factor levels considered in this design were foaming and non-foaming anaerobic digesters.

The procedure used to collect data for each of the variables is described below:

3.2.2.1 VS, TS, Alkalinity

These variables were determined using the methods as stated in Standard methods for the examination of water and wastewater (APHA, 2012) and expressed as alkalinity (mg/l CaCO3), total solid (%TS) and volatile solid (%VS).

3.2.2.2 pH

Measurement of pH was achieved using pH probe that was always calibrated prior to commencing the pH measurement.

3.2.2.3 Soluble protein

The soluble protein concentration in the samples was determined by the Bradford assay (Bradford, 1976) using Coomassie Brilliant Blue. The suitability of this method compared to other assays is its little interference from polyphenols (Siebert and Lynn, 2005). The method responds better to proteins greater than 3 kDa and to peptides comprising of aromatic (phenylalanine, tyrosine and tryptophan) and basic (arginine, histidine and lysine) amino acids.

Bradford reagent was prepared by dissolving 100 mg of Coomassie Brilliant Blue G-250 (Fisher, BP100-25) in 50 ml 95% ethanol in a 1 L volumetric flask plus 100 ml of 85% phosphoric acid and made up to 1 L with distilled deionised water. The reagent is stored in fridge and protected from light. Some precipitate may form in the reagent; this can be dissolved by gently inverting the bottle. Prior to analysis, the required volume of reagent were transferred to a beaker and allowed to reach room temperature before use.

Bovine serum albumin (200 mg/ml protein standard; Fluka, P5369) was used to prepare the calibration standards as shown in table 3-1. Diluted Bovine serum albumin (2 mg/ml in 0.15 M NaCl) was stored as 1ml aliquots in Eppendorf at 20^0C in NS1.30 and was used to prepare 100 µl standards as shown in table 3-2.

Samples and standards are analysed by adding 30 µl to a 2 ml Eppendorf and adding 1.5 ml Bradford reagent. The mixture is vortexed, and the absorbance read at 595 nm between 2 min and 1 hour. All samples and standards were analysed in triplicate. A typical calibration curve is provided in figure 3-3. The equation relating the absorbance (ABS) recorded and the concentration for each mixture prepared based on table 3-2 is as shown in equation 3-1:

Table 3-2: BSA calibration standards for Bradford assay

Standard [BSA] mg/l	BSA stock (µl)	Water (µl)
0	0	100
100	10	90
250	25	75
500	50	50
750	75	25
1000	100	0

Protein concentration (mg/l) = 0.7362(ABS)-0.0005 Equation 3-1

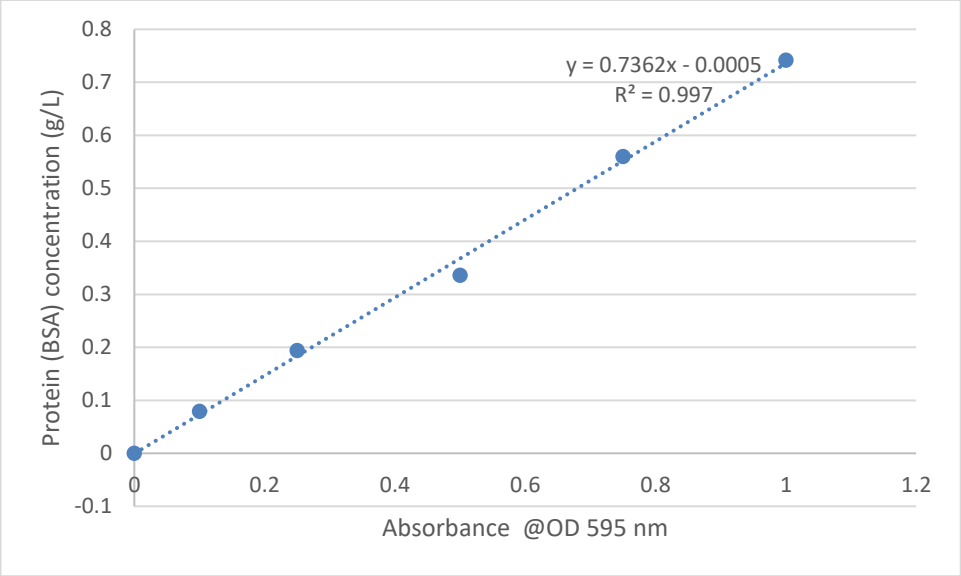


Figure 3-2: Typical Bradford essay calibration curve

3.2.2.4 **Volatile fatty acid**

Volatile fatty acid concentration (VFA) was determined using the Montgomery method (Montgomery et al., 1962). Montgomery et al., (1962) proposed an empirical method of determining organic acids based on

colorimetric ferric hydroxamate. The sample was heated with ethylene glycol and sulphuric acid, and the resulting mixture of esters was allowed to react with hydroxylamine. The hydroxamic acids formed were converted to their ferric complexes and determined by optical-density measurements at 495nm. Under the experimental conditions, the most intense colours were given by the volatile acids. The procedure as applied in this experiment is to;

1. Centrifuge 1.5 ml of sample in a microfuge tube at 6000 rpm for 10 minutes.

2. Carefully remove 0.5 ml of supernatant and add it to a 30-mL test tube.

3. Add 1.5 ml of ethylene glycol reagent and 0.2 ml of 19.5 N sulphuric acid then vortex.

4. Heat in a boiling water for 3 minutes and cool immediately in cold water, add 0.5 mL of 10% hydroxylamine hydrochloride solution and 2 mL of 4.5 N NaOH solutions then vortex.

5. Add 10ml of 10% ferric chloride solution followed by 10mL of deionised water. Set aside for 5 minutes without stopper (to facilitate the escape of dissolved gases) then measure the absorbance.

6. Take the reading within 1 h of colour development. If the concentration of volatile acid is expected to exceed 5000 ppm, set aside for 1 minute before adding ferric chloride and deionised water. The result of the VFA analysis is recorded as mg/L of acetic acid which is determined from an equation generated from the calibrated curve of known concentration of acetic acid against their specific absorbance.

Typical calibration curves as shown in figure 3-4 were produced and updated once new reagents were prepared as stocks get used up.

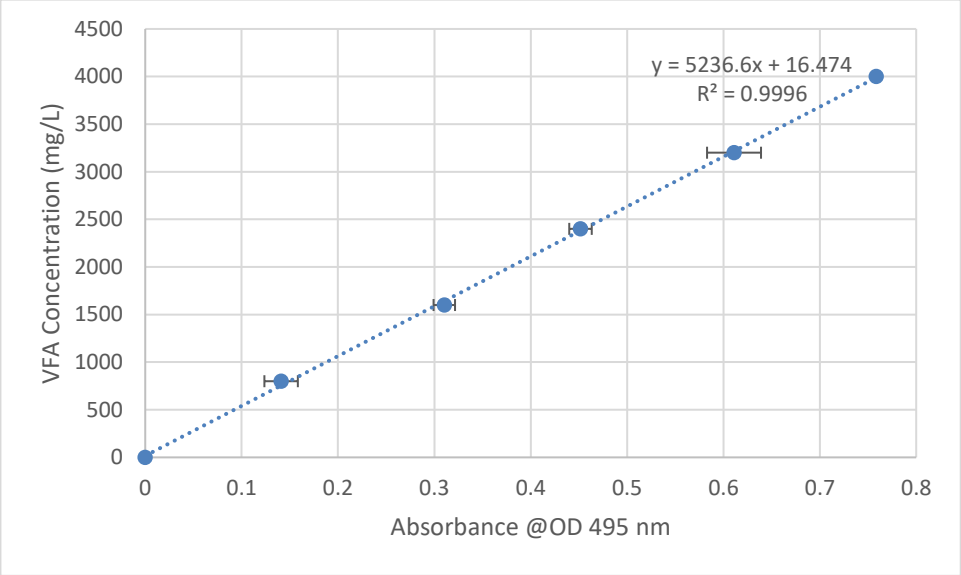


Figure 3-3: Typical calibration curve for determination of VFA concentration

3.2.2.5 Filamentous Index

Filamentous bacteria identification was performed using bright field microscopy and Confocal laser scanning microscopy together with staining methods such as gram stain fluorescent stain as discussed in section 2.6.6 and 2.7.1.1. This method differentiates between the Gram +ve or -ve bacteria depending on the existence of a permeability barrier in bacteria based on the chemical and physical properties of the cell wall (Mesquita, et al., 2013). Most bacteria in ASP are Gram positive with the foaming related ones notably Norcadia forms and Microthrix parvicella (Eikelboom, 2000). Using the pictures developed by EIkelboom (2000) and relating it to microscopic image of Gram stained sample, a value in the range of 1-5 was assigned to reflect filamentous index of the sample under study.

3.2.2.6 Foaming potential test

Foaming potential test was carried out using Alka-sletzer method. Two tablets of Alka-seltzer placed in a metal cage were introduced into 500ml graduated cylinder containing 250ml sample. The foam formation and collapse were monitored and recorded. The highest volume of foam within this period was recorded as the foaming potential. This was essential to assessing the propensity to foaming for the different sludge samples collected to ascertain

the impact of the operational condition and feed sludge composition on the foaming occurrence.

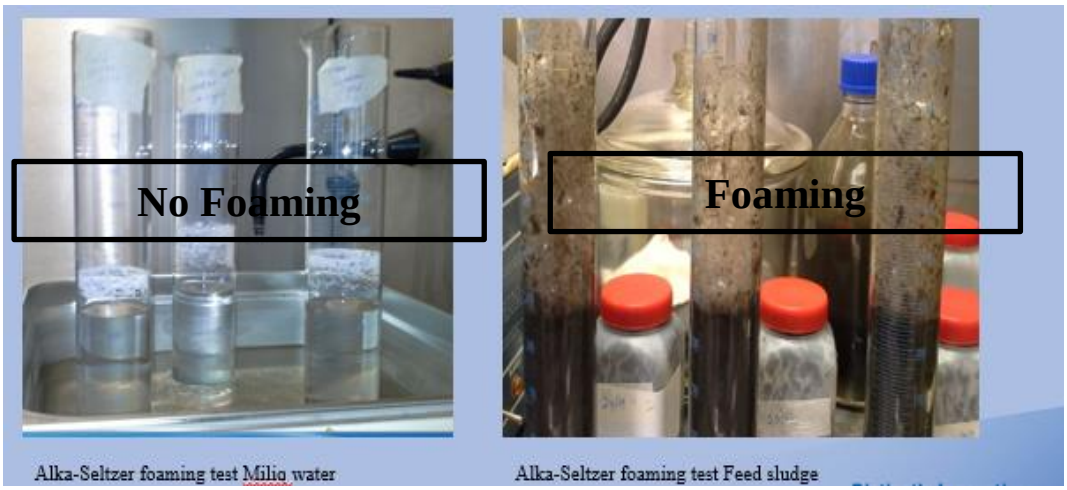


Figure 3-4: Alka-Seltzer® foaming test

3.2.2.7 Surfactant concentration

Anionic, cationic, and non-ionic surfactant were determined using Hach cuvettes; LCK 433_Anionic Surfactants, 0.05-2.0 mg/l, LCK331_Cationic surfactants, 0.2-2.0 mg/l and LCK 332_Non-ionic surfactants, 0.2-6.0 mg/l. The procedure as applied in this experiment is to centrifuge 1.5 ml of sample in a microfuge tube at 6000 rpm for 10 minutes. The supernatant carefully removed and used based on the volume stated on the different Hach cuvettes manual. This was carried to evaluate the impact of surfactant concentration on foaming occurrence.

3.2.3 Laboratory experiment to generate modelling data

3.2.3.1 Experimental design

Sequel to knowledge gained from literature review, historical data/visit to full scale AD and preliminary experiment on monitoring foaming in samples taken from full scale AD, experimental design comparing two treatments of foaming and non-foaming AD was proposed for this study.

The experiment involves using digester feed comprising a mixture of 60% primary sludge and 40% Secondary sludge. This ratio was reported by Eddy/AECOM (2014) as the most widely used ration for anaerobic digestion

of a mixture of primary and secondary sludge in most WWTW using AD. This ratio was also supported by the result of the survey that was carried out for WWTW in Scotland using AD.

The digestion process took place in 6, 1 L, glass bottles (Fisher Scientific Ltd., Loughborough UK) run as triplicates for the foaming and non-foaming AD. The digesters were placed in a thermostatically-controlled water bath (Patterson Scientific Ltd., Luton, UK) at 35 ^{O}C operating at a 500-mL digester volume to provide sufficient space to monitor the foaming occurrence as shown in figure 3-5. The digestion bottles were sealed with rubber bungs (Fisher Scientific Ltd., Loughborough, UK) and gas was collected in glass collection columns (length: 110 cm, diameter: 5 cm) by displacement of water containing hydrochloric acid (pH <2).

The samples collected from the digester during the feed were further analysed to establish data for the predictive model. In addition, maximum volume of foam was measured, and the digester shaken before the gas reading is taken. Gas samples were collected from the digester for analysis. Close to the end of the experiment, sludge samples were collected from foaming and non-foaming digesters and were used in the 'omics' analysis discussed in chapter 6.

Figure 3-5: Laboratory Anaerobic digester set up

Prior analysis and various researchers (Dalmau, et al., 2008; Ganidi, et al., 2011; Eskicioglu & Ghorbani, 2011, Subramanian & Pagilla, 2014) have highlighted organic overload as one of the major cause of AD foaming as discussed in section 2.6.3. Consequently, two levels of the factor OLR was used to permit the comparative study of foam formation due to variation in the digester OLR. In the two set of triplicate AD experiments, OLR was increased until the initial observation of foaming occurrence which was not nuisance but took place after feeding the digesters. Subsequently, increment in the digester OLR was ceased in the digesters marked to represent non-foaming (A, B and C) such that the same OLR at which initial foaming was observed was maintained for the rest of the experimental period. On the other hand, increment in OLR for the digesters marked to represent foaming (D, E, and F) was continuously increased for the rest of the experiment. The approach was channelled towards comparing the response of the monitored variables for a foaming and non-foaming digester in order to establish the variables that influences foam formation in AD process.

On daily basis, the digester was fed with equal amount of sample collected from a full-scale foaming AD treating activated sludge. The feeding was steadily increased by 10% every three days based on organic loading rate (kgVS/m3d). This was in accordance with recommendation in WEF (2007) as discussed in section 2.5.5.2. As the digesters were started off with active inoculum whose feedstock quality closely resemble the feedstock for the experiment and was gradually increased, the adaptation of the microorganism was faster resulting to full digester start up within 7days as proposed in WEF (2007). In addition, the ease of adaptation of the microorganism and reduction in lag phase of bacterial growth when using active inoculum and starting off with low volume of feed was discussed in section 2.5.3. The increment in OLR was carried out every three days in order to allow sufficient time for the digester to adjust to the change in the OLR before another increment.

Other experiments carried out during this study to widen knowledge on the impact of other factors on the foaming process in AD but not reported include;

1. Use of waste activated sludge (WAS) only. The same WAS that was used as a mix for experiment 1 was used as the only feedstock for this experiment 2.

2. Use of primary sludge (PS) only. The same PS that was used as a mix for experiment 1 was used as the only feedstock for this experiment 3.

3. Use of glucose and Triton X100 as feedstock to the digestion process was carried out to monitor the impact of surfactant on the AD foaming occurrence. Trition X100 addition served as an indication of increase in surfactant concentration.

The variables determined during the experiments and analytical methods used are discussed in the following sub-sections.

3.2.3.2 Foaming occurrence

It was identified that due to the intrinsic nature of the anaerobic digestion process and the persistent presence of surfactant material and gasses (Ganidi, et al., 2009), foaming was inevitable thus care was taken to identify and monitor nuisance foaming. The highest volume of foam observed in the digester was recorded. The digester which is a 1 L Duran flask was calibrated and measurement of the foam volume taken prior to feeding while foaming occurrence was monitored after feeding. The maximum volumes of foam observed in the digester were recorded as well as the volume of foam before the feed. This was crucial in identifying the stability of the foam.

Two calculations were carried out for the foaming occurrence; the maximum volume of foam as a percentage of digester volume and additional digester volume covered by foaming. The former was essential to reflect the total foaming occurrence in the digester (equation 3-4) while the later accounted for nuisance foaming which is the overflow that could occur when the digester foaming is enormous (equation 3-5) resulting to various complications.

$$Maximum\ volume\ of\ foam\ as\ a\ percentage\ of\ digester\ volume = \frac{Maximum\ foam\ volume}{Active\ digester\ volume}\ x100$$ Equation 3-2

$$Additional\ digester\ volume\ covered\ by\ foam\ as\ a\ percentage\ of\ digester\ volume = \frac{Maximum\ volume\ of foam - active\ digester\ volume\ (500\ mL)}{Active\ digester\ volume\ (500\ mL)}\ x\ 100$$ Equation 3-3

3.2.3.3 Organic loading rate

This was determined based on the volatile solid concentration of the feed sludge. Volatile solids and total solids were determined as stipulated in (APHA, 2012) and in equations 3-6 and equations 3-7. Attempt was made at comparing the VS concentration with chemical oxygen demand (COD) that was determined using Hach Lange cuvettes. However, feeding of the digester was carried out specifically based on the VS concentration. The values of TS and VS were applied to equation 3-2 to determine OLR.

$$TS = \frac{W_{total-}\ W_{dish}}{W_{sample-}\ W_{dish}} \times 100$$ Equation 3-4

Where;

$W_{total}\ = Weight\ of\ dried\ residue\ and\ dish\ \ (mg)$

W_{dish} = weight of dish (mg)

$W_{sample}\ = weight\ of\ wet\ sample\ and\ dish\ (mg)$

$$VS = \frac{W_{total-}\ W_{ash}}{W_{total\ -}\ W_{dish}} \times 100$$ Equation 3-5

Where;

$W_{ash} = weight\ of\ residue\ and\ dish\ after\ ignition\ (mg)$

3.2.3.4 Biogas production rate

In this study, biogas produced from the anaerobic digestion process was collected in a cylindrical glass as explained in the experimental design on section 3.2.3.1. The difference in height was recorded on daily basis as well as the daily room temperature and atmospheric pressure. The volume of biogas was calculated, and measures taken during the calculation to offset the effect of atmospheric pressure.

In most laboratory experiments, a common method of biogas collection is by liquid displacement as they are inexpensive, easy to set up/use, robust and capable of working for long periods without maintenance. Thus, the use of displacement gasometers was deployed for this experiment as shown in figure 3-6. Measurements taken directly from the gas column (e.g. liquid levels, pressure) were used to calculate gas volumes. Consistent measurement of room temperature and atmospheric pressure during gas measurement was routinely carried out to account for possible errors in volume calculations. As well as adjusting to STP for ease of comparison, it is essential to take into account the vapour content and to make a correction for any hydrostatic pressure on the gas (Walker, et al., 2009). Currently, there are several definitions of standard temperature and pressure (STP) in widespread use such as the International Union of Pure and Applied Chemistry (IUPAC) definition of 0 ^{o}C and 101.325 kPa and the National Bureau of Standards definition of 25 ^{o}C and 100 kPa. This variation could result to a difference of more than 10% for the same mass of gas (Walker, et al., 2009).

In this experiment, due to the design of the gas collection unit the following assumptions were made: the cross-sectional areas of the columns are constant; biogas acts as a perfect gas; once leaving the anaerobic digester biogas quickly cools to ambient temperature; the biogas is saturated with vapour but because of the length of the pipe, the vapour content of the biogas tend to condense in the pipe before reaching the collection column thus there is no

need to account for vapour pressure in the biogas leaving, the biogas pressure is the same as the atmospheric pressure in the room; the IUPAC definition was deployed in standardising the volume of the biogas.

Figure 3-6: Biogas collection set up

The volume of biogas produced and recorded under normal condition can be converted to STP using Combine Gas law as shown in equation 3-6 and 3-7:

$$\frac{P_n V_n}{T_n} = \frac{P_s V_s}{T_s}$$ Equation 3-6

Where;

P_n = normal atmospheric pressure (milibar)

V_n = Normal Volume (mm^3)

P_s = Standard pressure (1013.2 milibar)

V_s = Standard volume (mm^3)

T_s = Standard temperature (0^0C)

T_n = Normal temperature (0^0C)

$$V_s = \frac{P_n V_n T_s}{T_n P_s}$$ Equation 3-7

Unfortunately, gasometers stores gas and does not offer direct measurement of flow rates. Conversion of the standard biogas volume to flow rate (mm^3/hr) was essential to ensure consistency in analysis and avoid variations resulting from varying digester feed time. This was achieved by dividing the standard volume of gas by the difference in time recorded after the feed and prior to feeding the digester and collecting gas from the gasometer the next day as shown in equation 3-8.

$$BPR\ (^{mm^3}/_{hr}) = \frac{V_s}{t_2 - t_1}$$ Equation 3-8

Where;

$t_1 = time\ of\ initial\ feed$

$t_2 = time\ of\ next\ feed$

3.2.3.5 Composition of carbon dioxide in biogas

Monitoring the biogas composition in terms of the carbon dioxide and methane content was vital in assessing the performance of the AD (Kanu, et al., 2015a). This could be done using a gas chromatography equipped with a thermal conductivity detector (GCTCD) machine present in the laboratory. However, carrying out daily analysis of gas composition in replicates for a long period of time was not economical and viable for this research. Moreover, the GCTCD was not designed to be exposed to such extensive use.

Alternatively, gas scrubbing was adopted such that known volume of biogas was passed through an inverted cylindrical tube of known diameter in a 3M solution of NaOH, the difference in gas height in the liquid was used to calculate the volume of biogas remaining from the scrubbing. However, since the NAOH solution was meant to scrub out carbon dioxide present in the biogas sample, the difference between the initial volumes of biogas and the

volume occupied in the cylinder after the scrubbing accounted for volume of carbon dioxide present.

This was carried out at the same time for the same sample on the GCTCD, generating a calibration curve of the readings between the GCTCD and gas scrubbing with a correlation of 89%. This illustrated that the gas scrubbing was efficient in determining the carbon dioxide composition in the biogas sample. Thus, the gas scrubbing was eventually used for the rest of the experiment to determine the % of carbon dioxide present in the biogas.

3.2.3.6 Temperature

The temperature of the digester was maintained at 35^{O}C using a thermostatically set water-bath as shown in figure 3-5. A thermometer was placed in the water bath to monitor and ensure that the set temperature was maintained during the period of the experiment.

3.2.3.7 Volatile solid reduction (VSR)

As discussed in section 2.5.6.3, adopting Van Kleeck formular as in equation 2-4 may seem more useful as he assumed non-removal of supernatant. In addition, it was observed that the operators of full scale AD in the WWTP that supplied the historical data for this study were using the same formula as that was more representative of the prevailing digester operation.

3.2.3.8 Soluble protein, Volatile fatty acid and pH

The same procedure used in section 3.2.2.2, 3.2.2.3 and 3.2.2.4 was applied for the determination of soluble protein, volatile fatty acid and pH respectively.

3.2.4 Statistical analysis

The following statistical analysis was applied in the purpose of this research to assist in making decision and drawing conclusion from the data on various AD operations both at full scale and laboratory scale as the necessary.

3.2.4.1 Data pre-processing

In order to obtain reliable analytical results, data pre-processing is essential as it provides techniques on improving and validating measurement (Rustum & Adeloye, 2011). These involve identifying missing data and outliers and replacing them with 'NAN' (not a number) to comply with MATLAB and MINITAB requirements.

However, most of the result used in this study was experimental result thus there was absence of missing data. This was essential during the historical data analysis where lot of data were not available during digester recovery from system perturbation ranging in weeks. As an example, there was lack of data available for Hatton AD within the period of July to November 2014. This resulted in the use of NAN for the period to allow for the data to be analysed for the period that there was sufficient data. It is worth noting that, the basis for the historical data was to have a clear understanding of the on-going in the digesters. The approach would have been different if the data was to be used directly for the predictive model.

3.2.4.2 The sample size

It is vital that appropriate sample size is selected for any statistical analysis as it enhances the level of confidence in the statistical outcome. Yamane (1967) proposed equation 3-9 for the determination of sample size

$$n = \frac{N}{(1 + N\,(e)^{2})}$$ Equation 3-9

Where;

n = sample size,

N = Population size,

e = level of precision (0.05 at 95% confidence level.)

3.2.4.3 Descriptive statistics

The following descriptive analyses were chosen to enable an effective statistical analysis of the experimental results. All statistical analysis was carried out using Minitab 17. Mean and other measures of dispersion such as Standard deviation, Quartile, Maximum and Minimum Value were very useful in understanding the extent to which the sample data is spread out from the mean. Scatter plots, Interval plots and correlation analysis were deployed to highlight the relationship existing between the variables used in analysis in this study.

3.2.4.4 Normality test

As established in section 2.8.1.3.1, most statistical analysis is dependent on the data been a normal distribution. Hence, the test for normality was carried out using the absolute Z-score of skewness and kurtosis (equations 2-8 and 2-9) as proposed by Kim (2013) for sample data <50.

3.2.4.6 T-test

A t-test is one of the statistical test that is commonly used to determine whether the mean of a population significantly differs from the mean of another independent population. The default null hypothesis for a 2-sample t-test is that the two groups are equal while the alternative hypothesis is that the two groups are not equal. The test is calculated using equation 3-10.

$$\boldsymbol{t} = \frac{\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2}{\boldsymbol{s}/\sqrt{\boldsymbol{n}}}$$ Equation 3-10

Where: $\bar{x}_1 = mean\ of\ sample\ 1$

$\bar{x}_1 = mean\ of\ sample\ 2$

s = standard deviation of $\bar{x}_1 - \bar{x}_2$

n= sample size

3.2.4.5 Correlation analysis

Correlation analyses were carried out using MINITAB. The correlation analysis verifies the potential of any existing linear relationship. The correlation coefficient (r) signifies the strength of the relationship in the range of -1 to +1 representing negative or positive relationship respectively while 0 represent lack relationship between the two variables. The p-value explains the statistical significance of the identified relationship.

3.2.5 'Omics'study: samples preparation and data collection

The samples used for the 'omics' study were biological replicates taken over three days from the same reactor for the two classes of experiment namely; Foaming reactors before feed (PF), Foaming reactor 2 hours after feed (F), Non-Foaming reactors before feed (PNF) and Non-foaming reactor 2 hours after feed (NF), plus samples of Inoculum (I) and Feed (Fe). Collection of samples 2 hours after feed was based on the observation throughout the experiment that maximum foaming is achieved within the first two hours after feeding the digesters. It is worth mentioning that the digesters were fed semi-continuously which is once in a day.

The sample for the metabolomic and metagenomic analysis was collected at the same time from the foaming and non-foaming digester. The sample for the feed was collected from samples fed to the reactors on the days that the samples were collected from the foaming and non-foaming digesters while the inoculum was taken from the storage (4℃).

3.2.5.1 Sample preparation of DNA for Metagenomics analysis

The DNA extraction was carried out using the QIAamp Fast DNA Stool Mini Kit as it provides fast and easy purification of total DNA from fresh or frozen stool samples. The principle and procedure followed was as outlined by the producer with the following key features;

1. Purification requires no phenol–chloroform extraction or alcohol precipitation,

2. Since stool samples contain many compounds that can degrade DNA and inhibit downstream enzymatic reaction, these substances are removed by addition of InhibitEX Buffer, which is specially formulated to separate inhibitory substances from DNA in stool samples.

3. DNA was eluted in low-salt buffer and to be free of protein, nucleases, and other impurities or inhibitors.

4. The purified DNA was ready for use in enzymatic reactions, stored at –20℃ and transported to Glasgow Polyomics centre on dry ice for metagenomics analysis.

The simple QIAamp spin procedure yields pure DNA ready for direct use in as little as 25 minutes. The eluted DNA is up to 20 kb long and is suitable for direct use in any enzymatic reactions.

3.2.5.2 DNA Quantification

The DNA quantification was carried out using Nanodrop 2000c immediately after extraction before storage. The Nanodrop 2000c works by absorbance of UV-Visible light within 5 seconds and could be used for a concentration in the range of 0.4 – 15,000 nanograms/μL. In addition to determining the concentration range of the DNA, the spectral data and purity ratios indicate sample purity as no specific sample preparation is necessary. The DNA quantification was carried out following the procedure as outlined by the manufacturer. The samples were then stored at -20℃

3.2.5.3 Sample preparation for Metabolomic analysis

The sample preparation was based on the procedure recommended by Glasgow Polyomics center as outlined below for tissue extraction.

- Suspend tissue in 200µL or equivalent by tissue volume of Chloroform/Methanol/Water (1:3:1 ratio) at 4℃ or colder. Equal amount of 5µL pellet per 200µL of extraction solvent were measured for each sample to ensure consistency in comparison.

- The tissue was disrupted to lyse the cells by sonication then vortexed on cooled (4℃) mixer for 1 hour. This is essential in releasing the metabolites as tissues are resistant to lysis.
- The sample was the centrifuged for 3 minutes at 13,000g at 4℃.
- The supernatant (180µL) was taken and stored at -80℃ until analysis by LC–MS.
- A pooled sample containing approximately 5–10µL for each sample was prepared immediately after each sample extraction and used as a quality control sample in the LC–MS procedure.
- Since metabolites vary rapidly with change in the environment, processing of all sample replicates were carried out under identical conditions.
- Samples of the extraction solvent were also analyzed to allow removal of contaminants at the data-analysis stage.

The samples collected after the process were placed in screw-capped vials, stored at -80℃ and transported to Glasgow Polyomics in dry ice.

3.2.5.4 Metabolomics sample analysis

Hydrophilic interaction liquid chromatography (HILIC) was carried out on a Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific, Hemel Hempstead, UK) using a ZIC-pHILIC column (150 mm × 4.6 mm, 5 µm column, Merck Sequant)

The column was maintained at 30°C and samples were eluted with a linear gradient (20 mM ammonium carbonate in water, A and acetonitrile, B) over 26 min at a flow rate of 0.3 ml/min as follows:

Table 3-3: Metabolic sample elution

Time / minutes	%A	%B
0	20	80
15	80	20
15	95	5
17	95	5
17	20	80
24	20	80

The injection volume was 10 µl and samples were maintained at $4°C$ prior to injection. For the MS analysis, a Thermo Orbitrap Exactive (Thermo Fisher Scientific) was operated in polarity switching mode and the MS settings were as follows: Resolution 50,000, AGC 106, m/z range 70–1400, sheath gas 40, Auxiliary gas 5, sweep gas 1, Probe temperature 150°C and Capillary temperature $275°C$.

For positive mode ionisation: source voltage +4.5 kV, capillary voltage +50 V, tube voltage +70 kV, skimmer voltage +20 V. For negative mode ionisation: source voltage-3.5 kV, capillary voltage-50 V, tube voltage-70 V, skimmer voltage-20 V.

Mass calibration was performed for each polarity immediately prior to each analysis batch. The calibration mass range was extended to cover small metabolites by inclusion of low-mass contaminants with the standard Thermo calmix masses (below *m/z* 1400), $C_2H_6NO_2$ for positive ion electrospray ionisation (PIESI) mode (*m/z* 76.0393) and $C_3H_5O_3$ for negative ion electospray ionisation (NIESI) mode (*m/z* 89.0244). To enhance calibration stability, lock-mass correction was also applied to each analytical run using these ubiquitous low-mass contaminants.

3.2.5.5 Metagenomics sample analysis (16S Library Preparation and Sequencing)

16S library preparation and sequencing were performed by Glasgow Polyomics, using an adapted protocol based on the protocol provided by Illumina (San Diego, CA, USA). In brief, purified DNA samples, containing bacterial DNA of interest, were used as template to produce an amplicon library from the V3/V4 regions of the bacterial 16S ribosomal subunit. This amplicon library was then further amplified with Illumina Nextera XT adapters and indexes (San Diego, CA, USA) to allow sample clustering and multiplexing. All samples were normalised and combined in equimolar ratios, then run simultaneously on an Illumina MiSeq (San Diego, CA, USA), in the presence of 5% PhiX, utilising a 600 cycle (2x300 cycle) kit.

3.2.6 'Omics' Data analysis

3.2.6.1 Principle component analysis (PCA)

Visualizing data permit an easier understanding of the system under study using simple tool which include inter alia; boxplots, scatter plots, confidence interval plots and line plots. However, when large number of variables are involved such as in 'omics' study, the feasibility of using these univariate variables become extremely challenging demanding that the variables should be reduced while still exhibiting the typical variation of the complete data set.

Jolliffe (2002) identified PCA as the key tool for achieving reduction in large variable analysis. The low-dimensional scores are plotted as visualization/ pattern recognition tools, used as input in conventional classification, clustering and regression methods. Extensive explanation on the mathematical details of PCA can be found in Haykin (2009). Overall, PCA can be defined mathematically as a method that involves the discovery of the singular decomposition (eigen decomposition) of the covariance matrix of the sample under study.

In 'omics' applications, PCA is ubiquitously utilized to visualize data of high dimensional variables using two-dimensional plots of the principal

component scores (PCS) in order to ascertain the main patterns of variability as well as explore relationships (Ringner, 2008). Thus, PCS represent the uncorrelated combination of variables that articulate the most variance thereby highlighting the difference between the populations for the individual samples that make up the population (Hotelling, 1933).

3.2.6.2 Metagenomics

Quantitative insight into Microbial Ecology (QIIME) was used to perform microbial community analysis and interpret nucleic acid sequence data from the bacterial, and archaeal communities. To achieve this, multiplexed sequencing reads from high throughput 16S rRNA was processed through QIIME to generate taxonomic and phylogenetic profiles and comparisons of the sample presented. Principal component analysis, Chao, and Shannon diversity plots.

The Chao and Shannon diversity plots are very useful in understanding the result of the 16S rRNA sequencing and form part of the QIIME package. The Chao diversity plot is based on the number of rare OTUs found in a sample and very vital for estimating the richness of the species present in the sample (Chao, 1984) as show in equation 3-11:

$$S_{est} = S_{obs} + \left(\frac{f_1^2}{2f_2}\right) \qquad \text{Equation 3-11}$$

Where

S_{est} = estimated number of species,

S_{obs} = is the observed number of species,

f_1 = the number taxa represented by a single read in that community

f_2 = the number taxa represented by more than a single read in that community

If a sample contains many single reads of OTU, it is likely that more undetected OTUs exist, and the Chao 1 index will estimate greater species

richness than it would for a sample without rare OTUs. On the other hand, the Shannon diversity plot is a community diversity index that combines species richness and abundance into a single value of evenness. The evenness value is higher for microbial communities with equal distribution of abundance and vice versa (Shannon, 1948).

3.2.6.3 Metabolomics

Instrument raw files were converted to positive and negative ionisation mode mzXML files. These files were then analysed using the XCMS/MZMatch/IDEOM pipeline to produce the IDEOM file associated with this report. References for this are supplied below. Additional data output containing the complete result of the analysis is attached in the appendix 4 as IDEOM_Ify_Kanu_results.xlsb. Principal component analysis was also carried out to assess the reproducibility of the samples.

In addition, data were collected through 'omics' analysis of inoculum, feed sludge, foaming and non-foaming laboratory AD before and after feeding the digesters. The 'omics' (metagenomics and metabolomics result) were used to validate the results obtained during the laboratory experiment as they are high throughput and data driven.

3.2.7 Synthetic data generation

To ensure that the data presented to train the model was robust and representative of the foam formation in AD process, synthetic data generation techniques were used to extend the data records. Specifically, Markov lag-one autoregressive model as discussed in section 2.8.2.3 was applied using equation 2-7a, 2-7b and 2-7c.

The parameters determined from the experimental data include mean, standard deviation and lag one serial coefficient. These parameters were then applied in generating 3,100 data points. The initial value used in generating the synthetic data for each of the variable was the mean of the observed data, hence, the first 100 generated data was discarded as to allow the data be a better representative of the observed data.

3.2.8 Predictive modelling of foam formation in AD

The predictive model was developed using MATLAB software. The following modelling algorithm as discussed in section 2.8.2.4 were selected and applied in this research based on their performance and to assess their suitability for predicting foam formation in AD;

1. Mamdani fuzzy logic model
2. Neural network fuzzy inference system (ANFIS in MATLAB)
3. Neural network-Levenberg Marqudart
4. Neural network-Bayesian regularisation
5. Neural network-Scaled conjugate

Adequate care was exercised in order to ensure that the data were prepared, presented to the network appropriately and training of the network was executed efficiently.

3.2.7.1 Models validation and performance evaluation criteria

The ability of the model usually determines the exactitude of a model to reproduce accurate outputs of the AD process given the same input for which it was trained. Consequently, sequel to model selection and training, the models were evaluated by carrying out some validation test. The following evaluation criteria were used in assessing the accuracy of the model developed.

1. The Mean absolute error (MAE) measures the mean error of the predictions. MAE expresses accuracy in the same units as the data, which helps conceptualize the amount of error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(|x_i - x_i'|)$$ Equation 3-12

Where $x_i = Actual\ value$

$x_i' = predicted\ value$

n = number of observations

2. The mean square error (MSE): A commonly-used measure of accuracy of fitted time series values. Outliers have a greater effect on MSE than on MAE. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2$$ Equations 3-13

3. The root means squared error (RMSE) measures the mean root squared error scaled by the standard deviation of the values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|x_i - x_i'|)}$$ Equations 3-14

4. The correlation coefficient (R) measures the degree of linear relationship between two variables. The correlation coefficient assumes a value between −1 and +1. If one variable tends to increase as the other decreases, the correlation coefficient is negative. Conversely, if the two variables tend to increase together the correlation coefficient is positive.

$$R = \frac{1}{(n-1)S_x S_{x'}}\sum_{i=1}^{n}(x_i - \bar{x})(x_i' - \bar{x'})$$ Equation 3-14

Where

$\bar{x} = Sample\ mean\ for\ the\ actual\ data$

$x' = Sample\ mean\ for\ predicted\ data$

$S_x = Standard\ deviation\ for\ actual\ data$

$S_{x'} = Standard\ deviation\ for\ predicted\ data$

Visualizing the performance of the model is an invaluable means to assessing the performance of the model through simple plots that compare the predictions made by the model with the actual data. It simply depicts the performance of the model during the training, testing and validation of the model.

3.3 Material

The materials used in this experiment include the following;

3.3.1 Questionnaire

Based on the literature review and identified causes of foaming in anaerobic digester process as discussed in section 2.6, a comprehensive survey form was developed and attached as appendix 1 in this study.

3.3.2 Primary and secondary sludge:

As discussed in section 2.1, sludge removed from the primary sedimentation tank passes through the picket fence thickener where it is further thickened to about 4.28% total dry solid (TDS). The waste activated sludge (WAS) flows to the centrifuge belt thickener from the secondary clarifier where it is concentrated to 5.72% DS. The sample for the PS and WAS were collected after gravity and centrifuge belt thickening respectively from a sample collection valve prior to reaching the holding tank using a 500-mL bottle.

3.3.3 Import sludge (IS)

The average value of import sludge from other WWTP and Septic tanks were varying between 0.5 and 6% DS, they were further thickened and placed in the imports sludge holding tank. Sample collection for the IS was through a valve from the flow to the holding tank using a 500-mL bottle.

3.3.4 Inoculum

The digestate from a non-foaming anaerobic digester that was not dosed with any antifoam was collected and used as the inoculum for starting the laboratory experiment. The inoculum was collected with a 10 L container through the sampling valve at the bottom of the continuously mixed digester.

All sludge samples collected from the full-scale AD were stored at a temperature of 4℃ prior to using it to start the experiment.

3.3.5 Glucose

Glucose is a monosaccharide with the molecular formula $C_6H_{12}O_6$ obtained by hydrolysis of carbohydrates and was supplied by Sigma Aldrich for this study.

3.3.6 Triton X100

Laboratory grade Triton X100 which is a widely used non-ionic surfactant was purchased from Sigma Aldric

3.4 Conclusion

This chapter outlined the various methods and material that were applied in this research. This will be followed by the presentation and discussion of the result obtained through the application of the different methods in the course of this study.

The method and material were applied in the following: Chapter 4 focused on discussing the historical data and preliminary experiment which formed the basis for selecting the most appropriate variable for predicting foaming in AD; this led to the experimental design and set up discussed in chapter 5 through which data for the predictive model was collected and analysed; Omics study carried out on the laboratory experiment was discussed in chapter 6 as Metagenomics and metabolomics analysis of laboratory sample was helpful in creating a wider understanding of the AD activities and clarification of some mystery surrounding AD foaming; In chapter 7, synthetic data generation was applied on the collected data to generate a more

robust data that was fed into the predictive model. The models were evaluated to select the model that was most effective for predicting foaming in AD.

Chapter 4 : Selecting the most appropriate variables for modelling foam formation in AD

4.1 Historical data

Request for operational data and questionnaire (appendix 1) sent to the nine WWTWs (with AD) returned feedback and AD performance log from only four of the WWTW, those of which were operated by Veolia waters. The summary of the feedback is attached as appendix 2 and major points highlighted on table 4-1. Out of the four respondents, three of the AD have witnessed foaming in the past (Seafield, Hatton and Allanfern). Consequently, anti-foam dose was a consistent practice in these digesters to mitigate foaming occurrence with the use of anti-foam adding to the overall cost of running the AD. Only Newbridge AD was reported as non-foaming as nuisance foam have not been witnessed since its operation thus was the one digester that was not dosed with any anti-foam.

4.1.1 Discussion on feedback from questionnaire

It could be observed from table 4-1 that the volume for foaming digesters (HA and SF) were much more than the others with the non-foaming digester being the lowest in volume. This could imply that the digester size impacts the ability of the digester operating conditions to be maintained. It also points to the fact that it any be more beneficial to have smaller digesters than constructing huge AD in order to optimise digester operation.

As regards the dry solid (DS) concentration, it was realised that for the greater percentage of the different sludge stream (PS, WAS, IS), the foaming digesters had higher %DS of: AF (8.1% DS), HA (7.9% DS) and SF (5.72% DS). On the other hand, the non-foaming AD had the lowest % DS (3.2) for the highest sludge stream. This signifies that %DS of the feed going into the AD have a lot of control on propensity for AD to foam. Moreover, the OLR was highest for the persistent foaming digesters of 2.8 and 2.9 $kgVS/m^3$ for HA and SF respectively while AF and NB were at OLR of 1.3 and 1.5 $kgVS/m^3$ respectively.

Table 4-1: Major characteristics of foaming and non-foaming anaerobic digesters that samples used in this study were collected.

Digester configuration	Allanfean (AF)	Hatton (HA)	Newbridge (NB)	Seafield (SF)
No of cylinders	1	2	2	6
Diameter (m)	14.45	14	12.6	15
Height (m)	12	14	12.5	13.84
Volume (m^3)/AD	1979	2155	1559	2446
Total volume of AD	3958	4310	3118	14676
Proportion of feed to Digester				
	88% of digester feed is centrifuge thickened mixed sludge at 8.1 %DS:	GBT thickened sludge at 7.19 DS:		
	55% SAS	48% secondary	15% Thickened secondary sludge at 5.74 DS	36% thickened at 5.72% DS secondary sludge + import (Gravity belt thickener)
	33% Primary	44% primary sludge	33% Thickened primary sludge at 4.48% DS	64% thickened at 4.28% DS primary sludge (Picked fence thickener)
	12% is diluted imported cake at 9.30% DS	8% mixed import	52% Import mixed secondary primary at 3.2% DS (assumption, DS may largely vary)	

Operating conditions in the AD	AF	Hatton	NB	SF
Frequency of feeding	10 min / 30 min	10 min / 30 min	10 min / 30 min	20 min / 2 hours
Solid retention time	37.1 days	16 d	19 days	14.6 d
Temperature	33.2	36	Dig 1: 40 / Dig 2: 39.7	35
pH	7.61	7.1	Dig 1: 7.28 / Dig 2: 7.3	7.2, (5.7 feed)
Alkalinity	9535 mgCaCO3/l(outlet)	3221 mgCaCO3/l	Dig 1: 4215 mgCACO3/l ; Dig 2: 3901 mgCACO3/l	5700 (ideally in the range of 2000 to 4000mg CaCO3/L)
Organic loading rate	1.37 kgVS/m3.d	2.9 kgVS/m3.d	1.5 kgVS/m3.d	2.8 kgVS/m3.d
Mixing	Gas	Gas	Mechanical impeller	Sludge recycling
VFA	250mg/L (ideal 50-300mg/L)	N	13 / 15.8 mg/ L	N
Foaming	Yes; 2008 - 2012	Yes, 2 foaming events ever	No	Frequent foaming
Adverse conditions	Intermittent loading	Load increased in digesters	No	SAS/primary sludge ratio (more foaming when high SAS ratio, >50%)
	possible foaming in the ASP			Foam in ASP
	bad temperature regulation			Outlet pipe causing headlosses (might be too small)
Biogas (Nm^3/d)	2837	5863	1975	23255
Biogas(methane)%	59.77	61	64.5	63

Furthermore, the composition of the feed sludge stream affects the digester resistance to foaming as reported by some researchers (Niekerk et al., 1987; Ross and Ellis 1992). It could be observed that the Non-foaming digester (NB) had the lowest feed sludge stream of WAS at 15% of the total sludge feed to the AD. The report of the survey from SF identified that one of the adverse conditions usually witnessed at this AD is the increase in system perturbation following an increase in the ratio of WAS to PS feed to the AD. In addition, with NB digester feed comprising mainly of IS and yet not foaming, this illustrates that the quality of the IS could have impact on the digester operation as NB AD is contain humus which is taking from Blackburn trickling filter and septic tank sewage from different locations. Hence, it could be possible that the composition of the IS (% DS or microbial diversity) delivered at the NB digester had enabled it to overcome most operational challenges that could have resulted in system perturbation.

It was only SF that had a different feed pattern compared to other AD while AF was the only AD with a higher SRT. Thus, it could be deduced that the feeding pattern has an impact on the digester operation. This is because microorganisms are good at adapting to a specific feed pattern and once this is not achieved, the system will likely undergo some perturbation. In addition, consistency in the chosen pattern of feeding could affect AD stability. Thus, with a long SRT as in AF which is about twice the SRT for the other digesters could be contributing to the AD instability. Eddy/AECOM (2014) suggested that SRT in the range of 10-28 days is ideal for completely mixed AD with SRT of 10 days to be typical for smooth operation of digesters operated at a temperature range of 35-40℃.

The temperature and pH for all the digesters were fairly the same. However, these were average values for each WWTW as they all have more than one operating digester. Thus, the rate at which the factors vary over time may have an impact on the digester stability which may not have been captured in the data.

The method of mixing has also been identified by Pagilla et al. (1997) to have effect on the susceptibility of the AD to foaming. As reflected in the survey

report, Pagilla et al. (1997) concluded that gas-mixed digesters are more susceptible to foaming than mechanically-mixed digesters. NB digester was the only AD that was using mechanical mixing while the foaming AD were using both gas mixing (HA and AF) and sludge recirculation (SF)

The VFA for the NB was very low while the VFA for the other digesters were very close to 300 mg/L which have been stipulated as the limit that is adequate for the effective operation of AD treating sewage sludge (Eddy/AECOM, 2014). This signifies that VFA concentration play a role in AD foaming.

The total volume of biogas produced in AD was highest for SF which is expected as it has the highest volume and number of AD. On the other hand, the percentage composition of methane in the digestate for the NB digester was the highest showing that there is higher methanogenic activity taking place in the NB digester. It also reflects that the digester is more stable compared to other digesters.

4.1.2 Discussion on foaming AD (SF) performance log

The absence of foaming record and lot of discrepancies in the data collected for each WWTW AD limited the potential to use statistical analysis to explore the relationship in the historical data. These discrepancies include the following: while some WWTW had record for VFA and Alkalinity, others do not have such record; in addition, there were period when due to the digesters break down, there were no available data hence limiting the ability to establish a fair comparison of the data collected. Thus, time series plot of the persistent foaming digester (SF) was executed using volume of biogas produced as a benchmark to monitor AD stability as shown in figures 4-1 to 4-3for different variables.

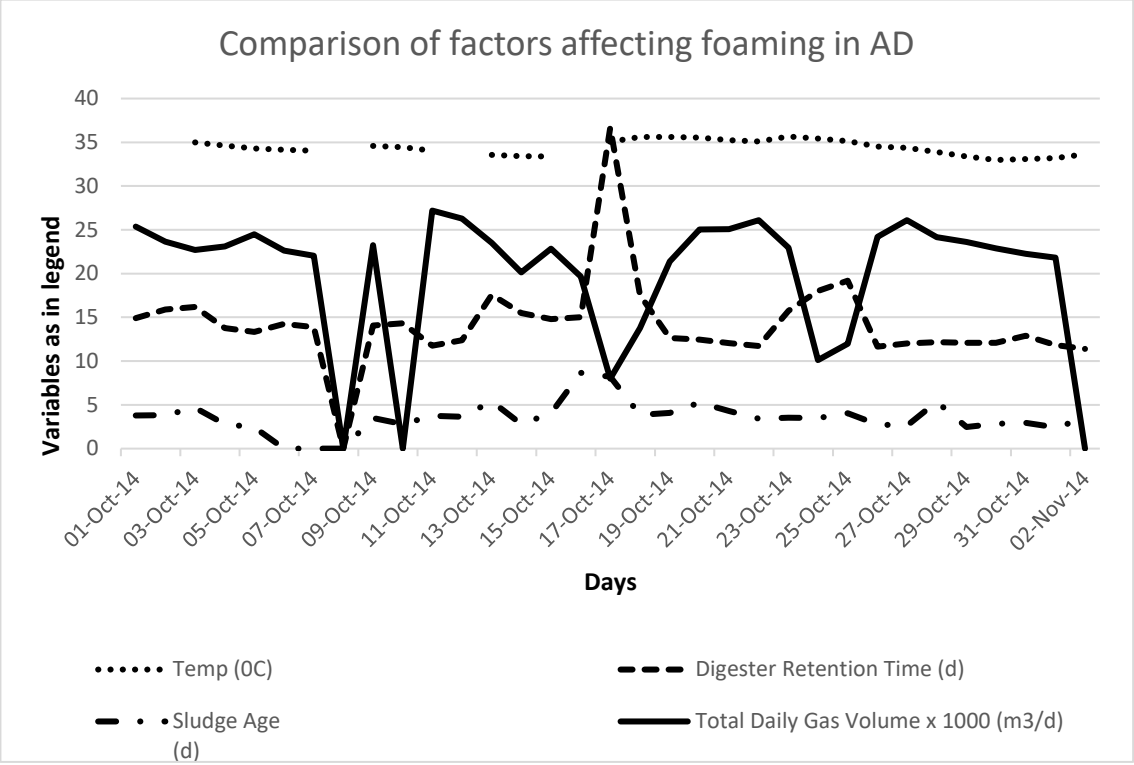


Figure 4-1: Relationship between factors affecting stability of AD

It could be observed from figure 4-1 that any variation in the aerator sludge age result to a proportional variation in the digester SRT which affect the composition of feed into the digester and the overall digester operation leading to system perturbation. Hence, significant variation whether an increase or decrease in sludge age and digester retention time resulted to a reduction in the volume of biogas produced by the digester. This is because a decrease in sludge age in the aerator result in more waste sludge being produced thereby increasing the amount of WAS relative to PS and vice versa when the sludge age is increased. As shown in figure 4-2, the period of lowest sludge age coincided with period of highest WAS generation

Variation in temperature of the digester ranged from 34℃ to 37℃. This is quite significant as a temperature variation of not more than 0.6 ℃ was ideal for efficient digester operation as reported by Eddy/AECOM (2014). Although the SCADA system is designed to ensure that temperature variations are properly managed, however, the huge size of these digesters make it difficult to achieve adequate circulation of sludge and temperature regulation.

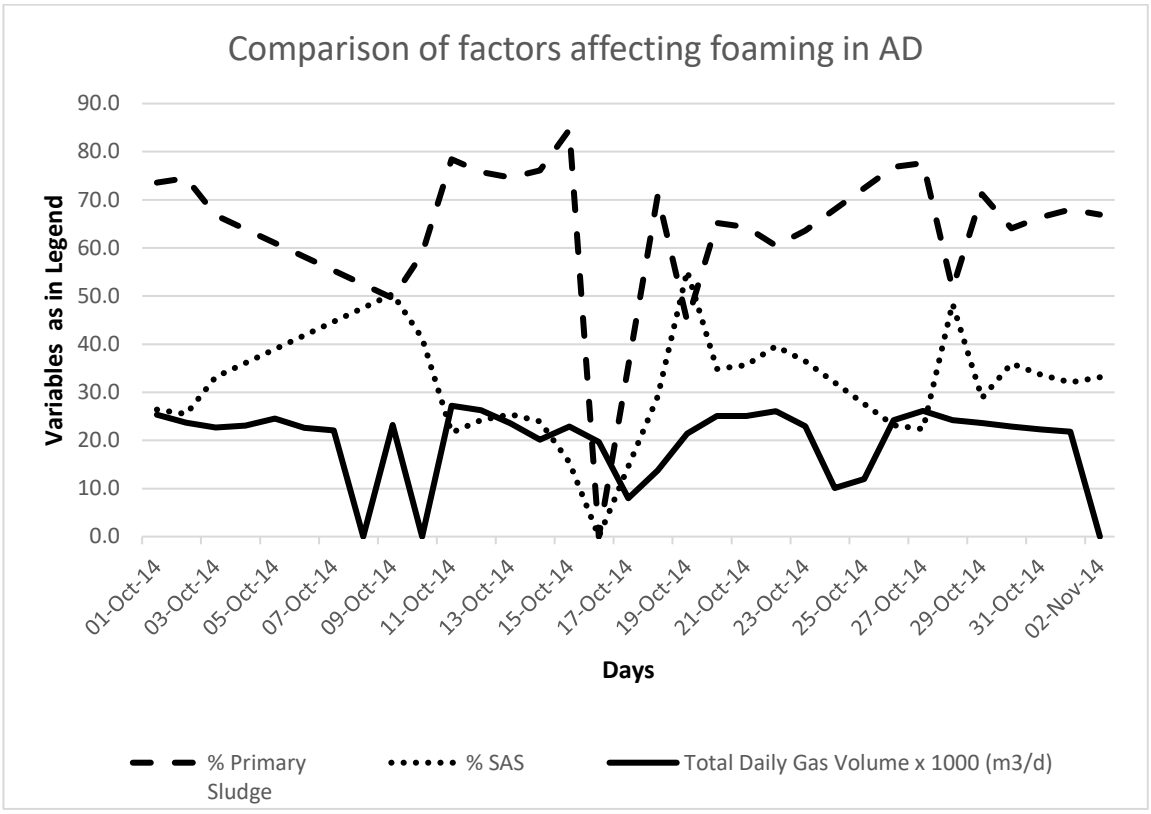


Figure 4-2: Relationship between factors affecting stability of AD

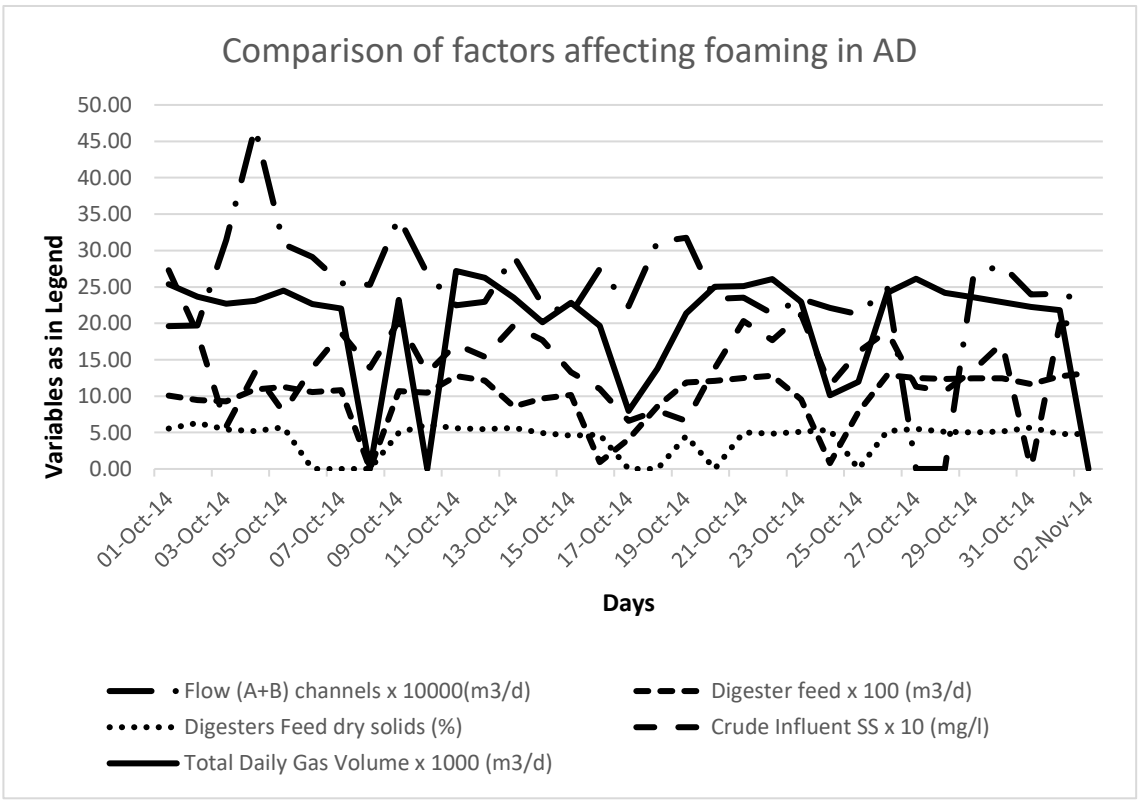


Figure 4-3: Relationship between factors affecting stability of AD

From figure 4-2, the variation in the composition of the sludge stream (WAS, PS) fed to the digester affect the biogas production as greater volume of biogas were produced during increment in PS compared to WAS, This support the discussion in section 2.1, that PS compared with WAS showed an enhanced degradability (Barber, 2014) thereby generating more biogas.

It could be seen from figure 4-3 that biogas production was obviously varying directly proportional to variation in the volume of feed sludge. This shows that an efficient control in the OLR is crucial to maintain the stability of the AD process resulting in greater consistency in the production of higher quality and quantity of biogas in the AD.

Following the discussions in this section, it is evident that the time series plot of a foaming AD was very helpful in understanding effect of the recorded operational variables on the propensity for foaming in AD. However, the use of antifoam and alkaline dose in the SF AD may have some impact on the data collected especially in terms of the volume of biogas produced. This is because of the antifoam/alkalinity dose cushioning off the effect of these variables on foam formation as well as allowing the digesters to keep functioning and producing biogas

4.1.3 Time series plot of Hatton AD

Hatton AD is unique in the extensive record of operational variables. In addition, it sparked interest as nuisance foaming occurred in the digester during the period of this study that led to a reduction in the digester feed and subsequent addition of nutrients to restore the digester operation. Thus it was also essential to review the time series plot for this AD in order to widen understanding on the key trend in AD operating variables prior to the foaming occurrence as shown in figures 4-4 to 4-7.

A close inspection of figure 4-4 highlights the lack of regularity in the feed quantity and quality. It could be observed that there is huge disparity in the diurnal volume of sludge fed to the AD which do not relate at all with the variation in feedstock quality in terms of the DS and VS composition.

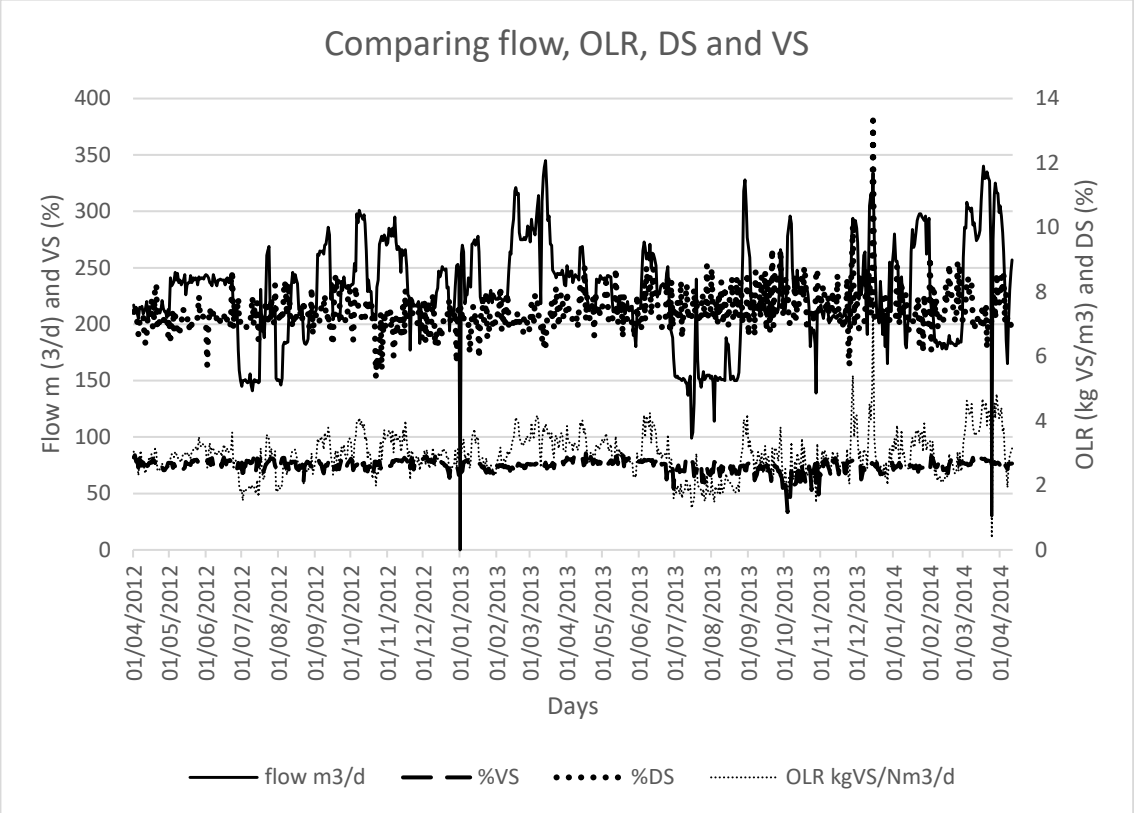


Figure 4-4: Comparing flow, OLR, DS and VS

The trend of the feed volume follows the same trend with the OLR which signify that OLR is controlled more by the volume of feed to the AD rather than the feed characteristics. This clearly demonstrate that the Programmable logic controller (PLC) of existing AD have been designed to increase the feed volume to the digesters based on the level reading in the feed sludge holding tank without due recourse to the feedstock characteristics and the impact it will have on the digestion process.

Unfortunately, microorganisms require time to adjust to different concentrations of feedstock. Complicating the AD process is the variety of microorganisms present in the digester resulting in different level of sensitivity to environmental changes amongst them as discussed in section 2.4. Hydrolytic and acidogenic bacteria are more resistant and proliferate more than the methanogenic bacteria, thus, the increase in the volume of sample with a resultant increase in the biodegradable content will result in higher rate of hydrolysis and acidogenesis, producing more acidic metabolites. This will lead to greater accumulation of VFA in the digester

making the digester more acidic and limiting the ability of the methanogenic bacteria to carry on with methanogenesis.

The overall impact of the above is the failure of the system to carry on with normal operation and may take some time for the digester to recover or in some instances may not recover at all without an external input such as alkaline dosing to buffer the system, seeding the digester with methanogenic bacteria to progress the conversion of the acids to methane or dosing with antifoam if foaming occurred.

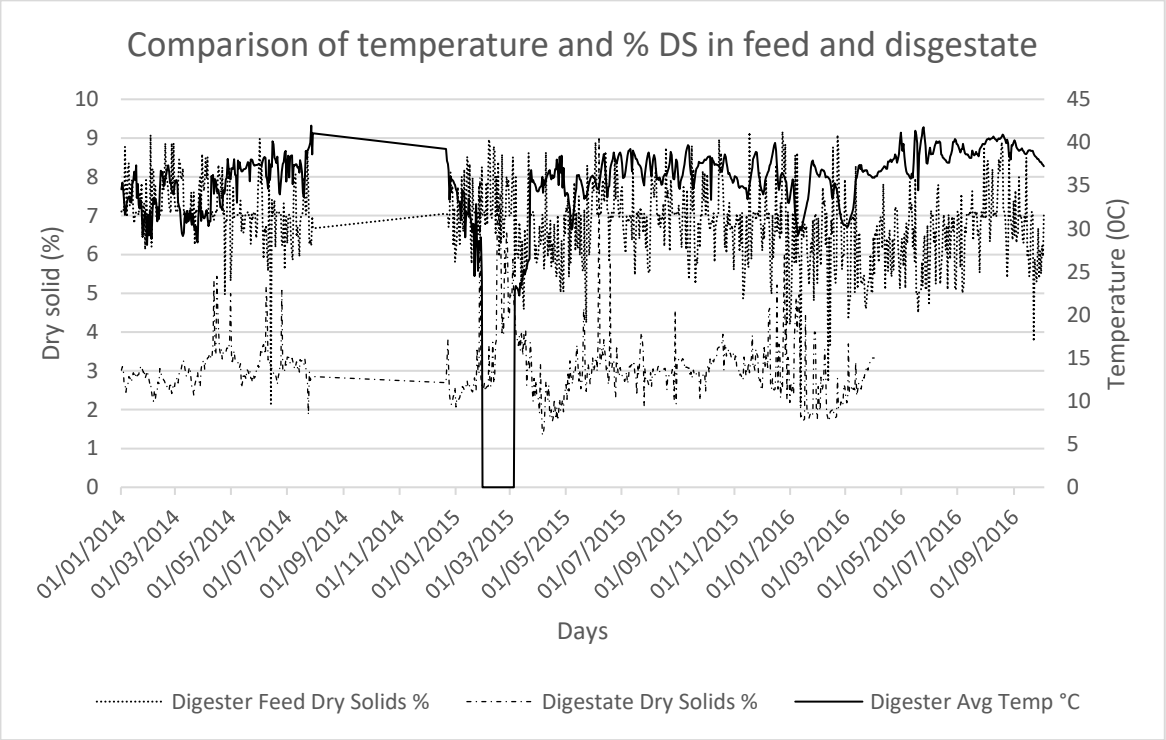


Figure 4-5: Comparison of composition of dry solids in feed and digestate

In figure 4-5, there is obviously no relationship between the digester and digestate DS concentration which further support the preceding discussion that there is no consistency in the feed quality. In addition, figure 4-5 depict a seasonal pattern in the month of March and July whereby there is continuous increase in the DS concentration in the digestate. This could be explained as the impact of temperature variations as we move from cold winter to spring and from spring to summer, the heat control may take some time to adjust to such weather changes, the resultant effect is that it will also affect the microbial activities coming into the wastewater stream and their subsequent

activities in the AD. In figure 4-5, it is obvious that it was a challenge maintaining the right temperature variation (±0.6℃) for the AD.

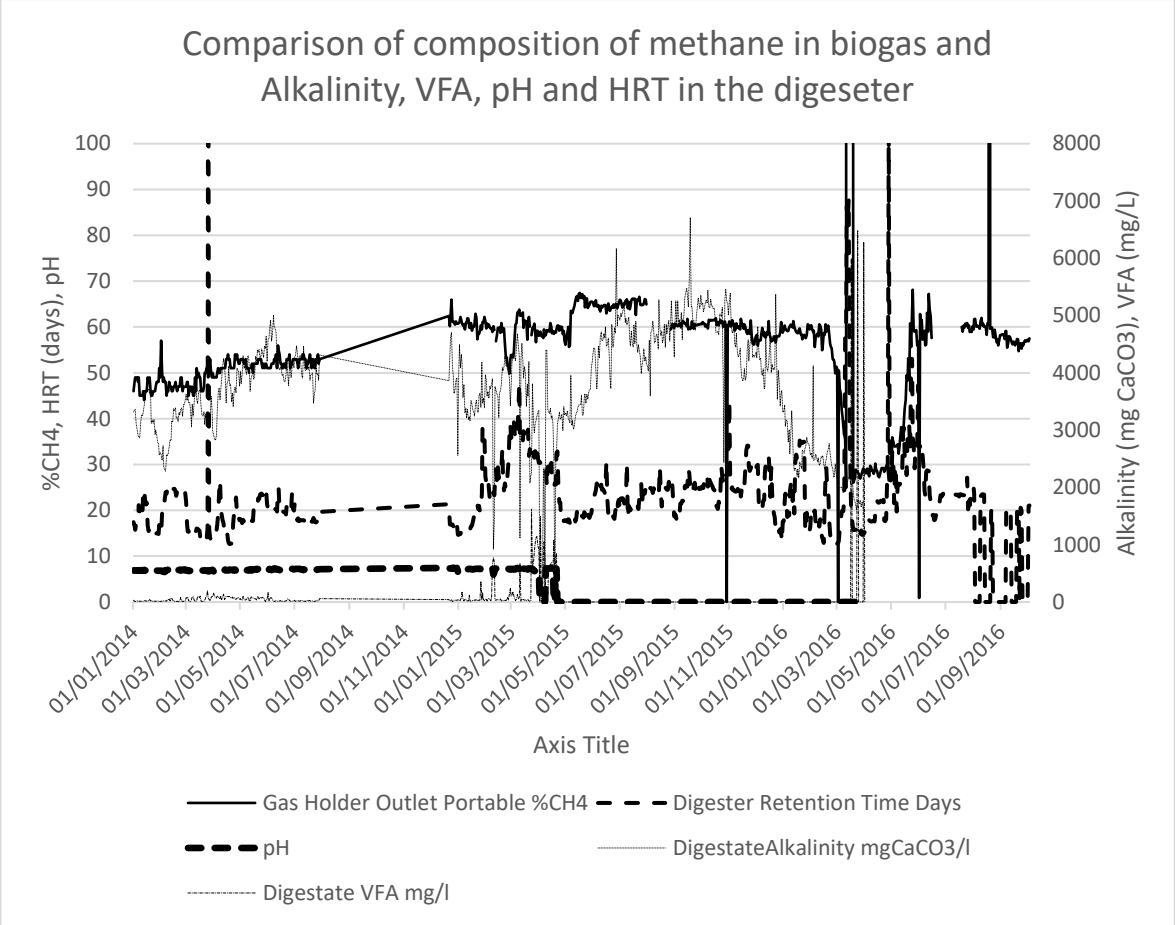


Figure 4-6: Comparison of composition of methane in biogas and alkalinity, VFA, pH and HRT in the digeseter.

The HRT of the digester was consistently varying which is expected due to the huge variation in the feed as shown in figures 4-4 and 4-5. The seasonal variation is also reflected in these variables as within the month of March, there were increases in the VFA, HRT and a decrease in alkalinity. pH and biogas production did not seem to respond quickly to these variations but did after some time as changes reflected by the other variables impact on the digestion process. It could be observed that during the last Hatton foaming episode which took place in April 2016, there were huge accumulation of VFA in the digester and a decline in alkalinity and methane production thus supporting the fact that methanogens are highly affected by adverse operating conditions such as VFA accumulation and lack of sufficient alkalinity to buffer the digester content.

VFA peaked on 24/03/2014 to 195.5, 24/03/2015 to 1616 mg/L and on 23/03/2016 to 6743 mg/L. It is worth pointing out that for 2016, there was no record for VFA from January 13 when VFA was last recorded as 20 mg/L until March 16, 2016 when it was recorded as 5761 mg/L.

The pH was recorded to be low during 05/ 02/2014, 9-11/02/2015 and no data for 2016. The drop in pH was related to drop in Alkalinity in the digester on 31/12/2014 and 9/02/2014. This drop-in alkalinity is connected to the increased VFA and the feed quality (%DS/VS). Thus, the digester buffering capacity was affected. This could also explain why the digester could not withstand the increment in VFA concentration due to increased organic loading rate at those periods.

Data on gas reading seemed steady most of the time and this is related to the fact that the readings were taken from the gas holder thus the gas had mixed up and are not very good representative of what is going on in the reactor at the point the measurements were taken. Notwithstanding that, the methane concentration in biogas varied directly with alkalinity and inversely proportional with VFA.

Eddy/AECOM (2014) identified that sludge that settles well have SSVI between 50 and 150. A continuous increase in SSVI from October 2015 showed that there were issues with the settleability of the SS. Peaks for SSVI occurred on 01/11/2015 and continued with the most significant peaks over 300 on 28/09/2016 and 09/08/2016.

Since SSVI is useful in measuring biological suspension from the activated sludge process, it serves as a tool to decipher the impact of hydrophobic substance on the digester operation. As noted in section 2.6.4, these hydrophobic substances may include: surfactants, biosurfactants and mycolic acid containing organism such as filamentous microorganism.

It is established in literature that sludge that do not settle properly usually are high in filamentous bacteria which have been identified as the microorganism that usually result in foaming occurrence due to their hydrophobicity. Thus, further analysis of the microbial and metabolic

constituent is essential to understanding whether the foaming occurrence is dependent on any of the notable hydrophobic substance found in AD.

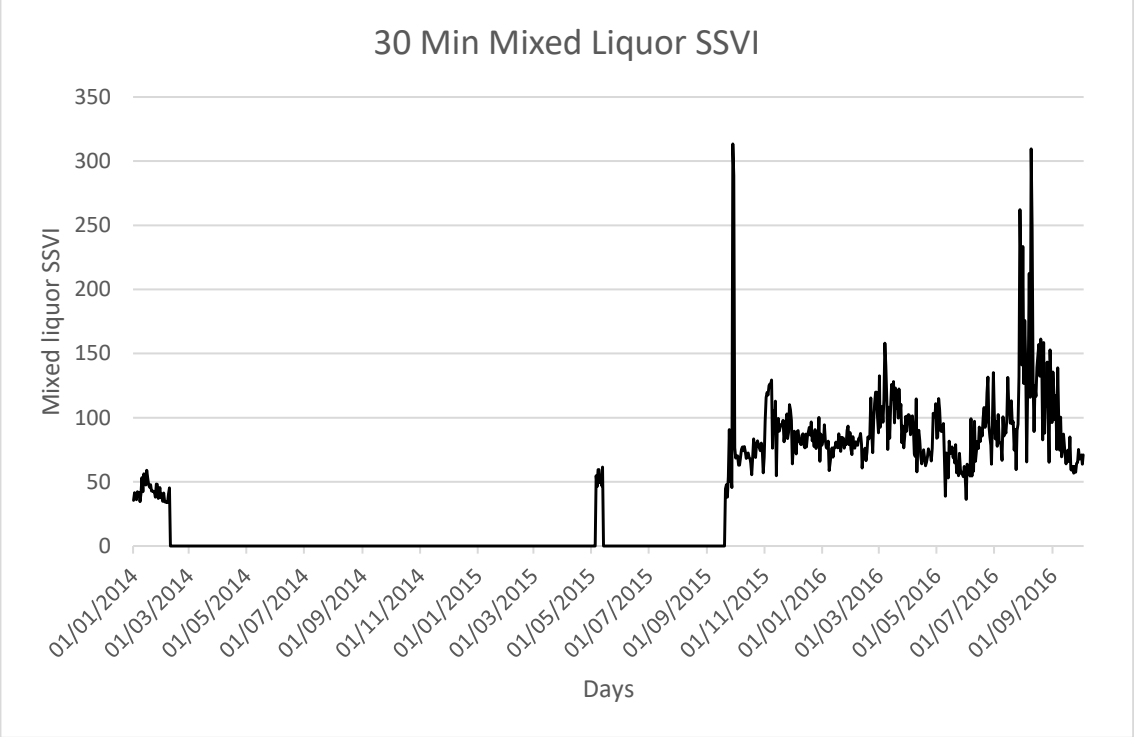


Figure 4-7 Hatton Digestate Mixed Liquor SSVI

As the purpose of studying the Hatton AD was to gain a further understanding of factors contributing to foaming in AD in a most recent foaming occurrence in April 2016, the following was deduced:

1.Foaming occurrence were caused by inconsistent feeding and temperature variation which was exacerbated by the presence of hydrophobic susbsatnce even though the digester was been dosed with antifoam.

2. The yearly/seasonal occurrence resulted in the lack of data for some months in 2014 and the same applied to some months in 2016. This is explained by the fact that consequent to the digester failure, the feed to the digester is ceased to allow it to recover or assisted to recover.

4.2 Preliminary experiment

The preliminary experiment was carried out as detailed in section 3.2.2 to extend the knowledge gained on AD operation and foam formation. This was

crucial due to the inconsistencies existing in data record for AD which often occur during digester failures.

Thus, the preliminary experiment was designed as a comparative study of the operating variables in AD (foaming and non-foaming) for feed and digestate samples. These operating variables include; pH, Alkalinity, VFA, DS, VS, FI, Foaming potential (FP), Soluble protein (SP), surfactant concentration.

4.2.1 Statistical analysis

Using Minitab, the t-test was applied to determine the variables whose variation are statistically significantly different between the foaming and non-foaming digesters both for the feed and digestate sludge in order to use them as the basic control factor in designing a more extensive study to monitor foaming occurrence and generate data for predictive modelling of foaming in AD.

4.2.1.1 Feed Sludge

The sludge samples were collected from Newbridge (non-foaming; NF) and Seafield (Foaming; F) full scale AD treating sewage sludge from their feed sludge holding tank where the mixture of PS, WAS and IS have been achieved. The results are summarised in table 4-2 to 4-3.

Table 4-2: Summary of descriptive statistics for feed sludge (Foaming and non-foaming AD)

Variables	Mean		SE Mean		Std Deviation	
	F	NF	F	NF	F	NF
pH	5.64	5.45	0.1	0.16	0.26	0.42
Alkalinity	1601	1225	143	300	377	795
SP	154.9	46	13.1	11.5	34.5	30,5

Surfactants	168.4	114.7	26.4	40.5	59.0	90.7
TS	5.25	3.23	0.63	1.51	0.24	0.57
VFA	1637	1570	235	192	621	507
FP	136.4	254.3	36.7	82.0	97.2	216.9
FI	2.43	2.57	0.65	0.30	1.72	0.79

Table 4-3: Summary of the t-test result for feed sludge (Foaming and non-foaming AD)

Variables	$\bar{x}_1 - \bar{x}_2$	**Confidence Interval (95%)**	**T-value**	**P-value**	**DF**	**Accept null hypothesis**
pH	0.184	-0.23, 0.6	0.99	0.34	10	Yes
Alkalinity	376	-391, 1143	1.13	0.29	8	Yes
SP	108.9	70.5, 147.2	6.25	0.00	11	No
Surfactants	53.7	64.7, 172.1	1.11	0.30	6	Yes
TS	2.02	0.59, 3.44	3.27	0.01	8	No
VFA	68	-600, 735	0.22	0.83	11	Yes
FP	-117.9	-325, 89.3	-1.31	0.23	8	Yes

FI	-0.143	-1.79, 1.5	-0.2	0.85	8	Yes

Where

$\bar{x}_1 = mean\ of\ foaming\ variable$

$\bar{x}_2 = mean\ of\ non-foaming\ variable$

The results of the descriptive statistics and t-test revealed that the differences between most of the variables are not statistically significant hence the null hypothesis that their means are equal was accepted. However, only TS and soluble protein depict that the difference in their mean are statistically significant to reject the null hypothesis that there mean are equal and there exist some differences between the foaming and non-foaming feed sludge samples for the variables TS and soluble protein.

4.2.1.2 Digestate sludge

The sludge samples were collected from Newbridge (non-foaming) and Seafield (Foaming) full scale AD treating sewage sludge from the the sampling valve located at the lower part of the digester and the result as shown in table 4-4 to 4-5.

Table 4-4: Summary of descriptive statistics for digested sludge (Foaming and non-foaming AD)

Variables	**Mean**		**SE Mean**		**Std Deviation**		**Variance**	
	F	NF	F	NF	F	NF	F	NF
pH	7.5	7.5	0.04	0.1	0.12	0.26	0.01	0.07
Alkalinity	7496	4694	329	505	871	1337	759048	1786498
SP	126.17	75.07	8.5	8.8	22.48	23.28	505.18	541.83

Surfactant	215.4	219.0	36.2	10	80.9	22.4	6542.3	503.7
TS	2.9	2.18	0.13	0.12	0.34	0.31	0.11	0.1
VFA	360	331.1	48.8	83.2	129	220.1	16641.7	48423.8
FP	404.3	330	33.2	32.7	87.9	80	7728.6	6400
FI	1	0	0.22	0	0.56	0	0.33	0

Table 4-5: Summary of the t-test result for digested sludge (Foaming and non-foaming AD)

Variables	$\bar{x}_1 - \bar{x}_2$	**Confidence Interval (95%)**	**T-value**	**P-value**	**DF**	**Accept null hypothesis**
pH	-0.001	-0.25,0.251	-0.01	0.99	8	Yes
Alkalinity	2803	1459,4146	4.65	0.001	10	No
SP	51.1	24.2,78.0	4.18	0.002	11	No
Surfactants	-3.6	-107,100.7	-0.09	0.929	4	Yes
TS	0.781	0.4, 1.16	4.51	0.001	11	No
VFA	28.9	-189.2,247	0.3	0.77	9	Yes
FP	74.3	-29.5,178.1	1.59	0.14	10	Yes
FI	T-test not calculated as all values in column were identical,					

As in the feed sludge, the results of the descriptive statistics and t-test revealed as shown in table 4-4 and 4-5 reveal that the difference between most of the variables are not statistically significant hence the null hypothesis that their means are equal was accepted. However, only alkalinity, TS, and soluble protein showed that the difference in their mean are statistically significant to reject the null hypothesis that there mean are equal and therefore accept the alternative hypothesis that there is existence of some differences between the foaming and non-foaming digestate sludge samples for the variables alkalinity, TS and soluble protein.

4.3 Conclusion

One significant observation during the experiment which is not captured in the statistical analysis is the display of very short-lived foaming period by NB samples which usually last within 4mins for the formation and collapse of foam though with a higher volume of foam. In contrast, SF samples had lower volume of foam but a longer foaming period. This could be attributed to the higher concentration of hydrophobic substances such as VFA, soluble protein and surfactants.

On the other hand, the result of this statistical analysis acknowledged the importance of TS and soluble protein as the key variables that could serve as a control for experimental design to monitor foaming AD. Although, alkalinity was relevant for the digestate sludge, the fact that it was not pertinent to feed sludge makes it less vital for predictive modelling. However, based on the historical data analysis, record of soluble protein was not part of the variables noted in full scale AD performance log. In addition, soluble protein does not only include the protein content of sample but also a makeup of the microorganisms present in AD, thus, it is not a true variable for controlling system perturbation, hence the need to focus on TS.

Since TS is not used directly in AD process but used to calculate the organic loading rate (OLR) fed to the digester, the OLR was selected as the key control variable in designing an experiment to monitor foaming in AD.

Furthermore, this analysis also identified OLR as the main cause of foaming in AD process. Sequel to the experimental design as detailed in section 3.2.2, data were collected and analysed in the next chapter.

Chapter 5 : Experiment to collect data for predictive modelling

Consequent to the variety of experiment carried out and the diverse data collected for each experiment, great focus will be on the first experiment based on which the data for modelling foaming occurrence were derived. The result of the other experiments as listed in 3.2.3.1 will be brought into the discussion.

Since the analysis carried out in section 4.2 identified OLR as the main causal factor for AD foaming, it formed the key variable that was altered in order to establish its impact on foaming. However, as other variables also contribute to foaming following the initial set up of the enabling environment due to variation in OLR, the following variables were also monitored during the experiment in addition to foaming include: tVFA in the feed and digestate, FI in the feed and digestate, percentage of carbon dioxide in the biogas, biogas production rate, pH, soluble protein in the digestate, alkalinity in the feed and digestate.

5.1 Discussion on result

Sequel to the experimental design, the three digesters were replicate of one another for each of experimental set, hence the following feed characteristics were the same:

- Volume of sample fed to the reactor
- Volatile solid concentration
- Organic loading rate
- Volatile fatty acid concentration (VFA)
- Filamentous Index

5.1.1 Volatile Solid Concentration

In order to ensure that the reactors were fed with the same quality and quantity of feedstock, the vessel holding the feed sludge was always mixed prior to collecting the feed sludge and a consistent volume of sample was maintained for each reactor.

Presented in figure 5-1 is the volatile solid concentration in the feed and the digestate showing a consistent but high volatile solid reduction in the digestate. However, the volatile solid concentration in the feed was constantly changing during the experimental period leading to changes in feed quality as the samples fed to the digester were collected weekly. Consequently, on day 13, the feed sample collected from WWTP was thicker and more viscous with higher volatile solid concentration leading to a sudden increase in volatile solid concentration by 46%.

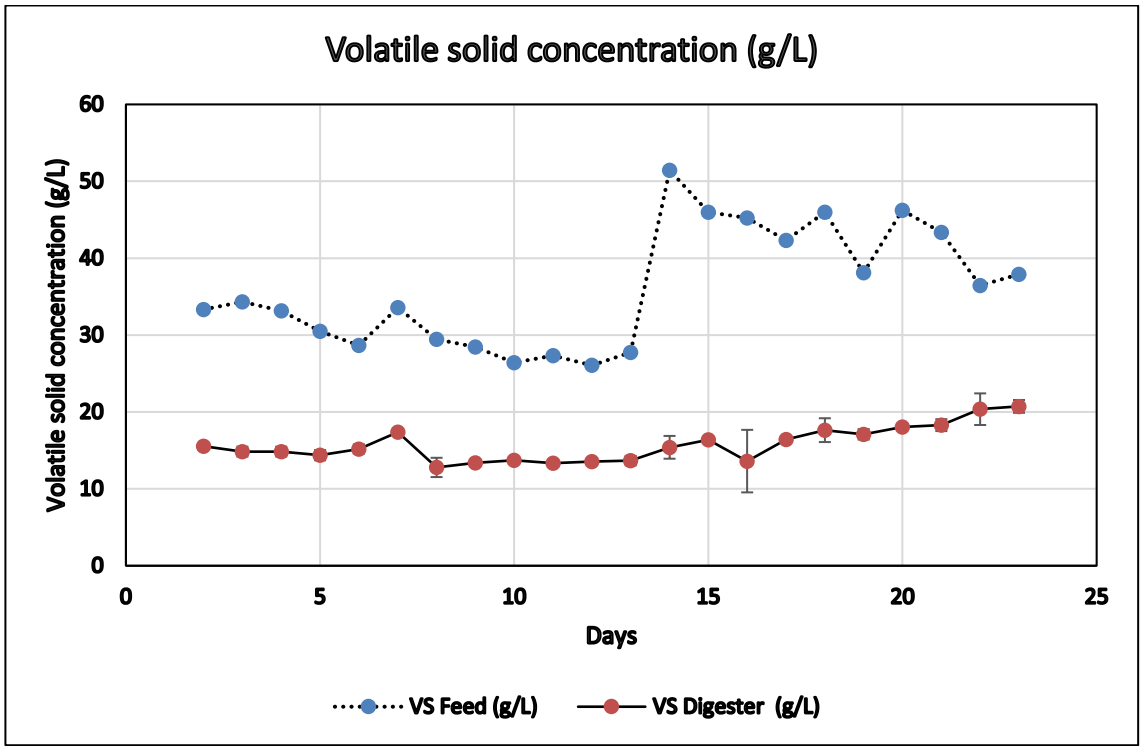


Figure 5-1: Variation in volatile solid concentration

This variation in the feedstock quality may have happened due to diurnal variation in the content and volume of flow to the treatment plant or because of changes in the operating condition of the different sects of wastewater treatment process.

Thus, variation in volatile solid concentration had a huge impact on the organic loading rate as shown in figure 5-2 which was varying differently irrespective of the consistent varied increment in the volume of feed sample as reported in figure 5-3. The change in OLR at day 13 is highly related to the changes observed for the volatile solid concentration (VS) in figure 5-1 which is expected due to the relationship between OLR and VS

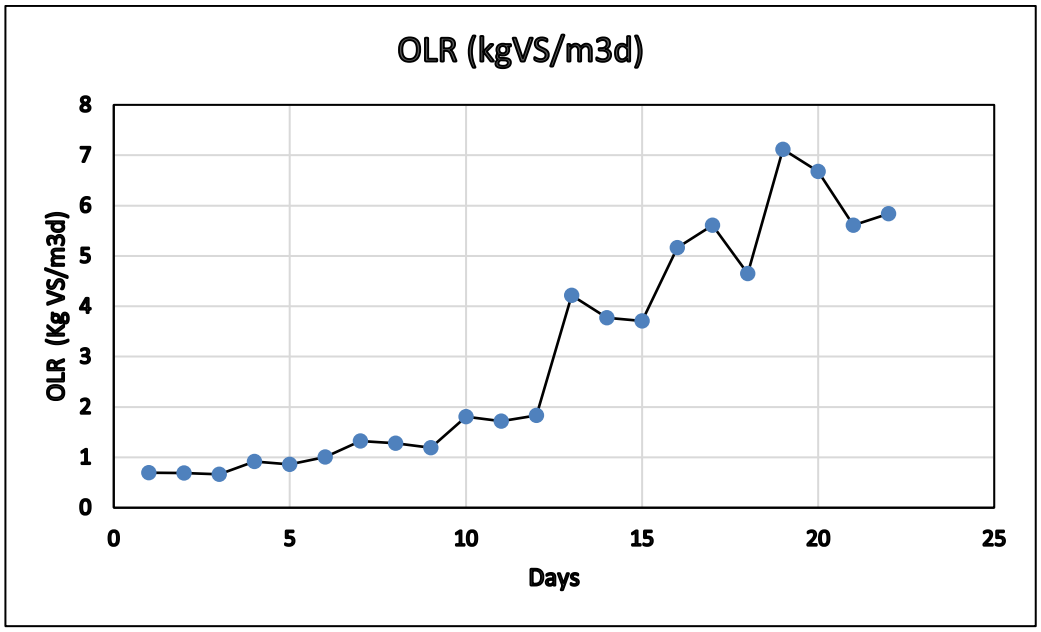


Figure 5-2: Variation in Organic loading rate

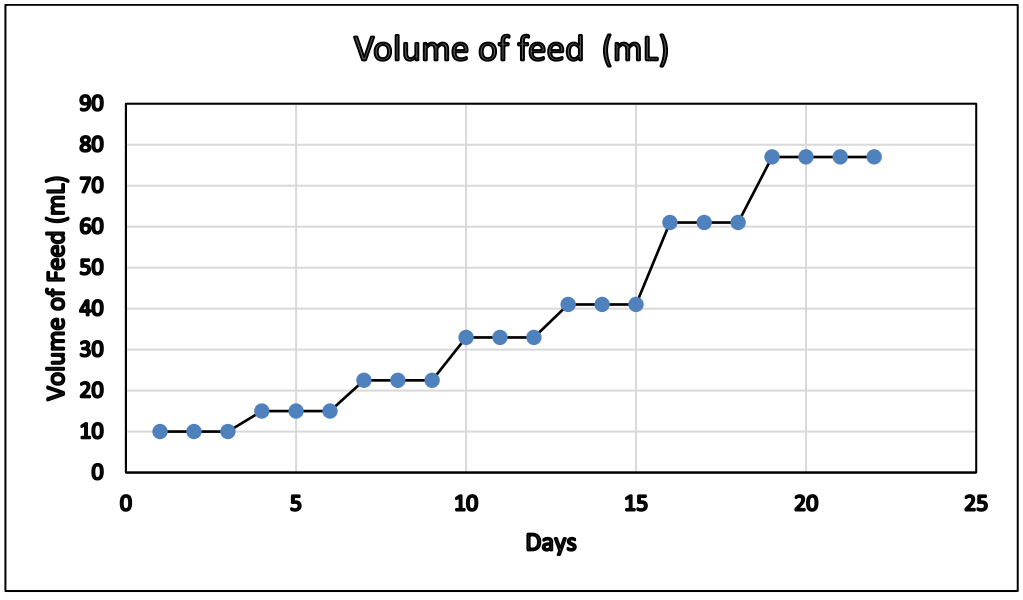


Figure 5-3: Variation in Volume of feed to the reactors

The results reported in figure 5-1, show that over the days, the VSR in the reactors continued to increase gradually showing that the reactors were no

longer performing optimally as it progresses from an optimal operating condition until day 20 when the digester began to go sour.

This is typically what goes on in most AD treating sewage sludge of which the feed to the digester is controlled by the volume of feed fed to the digester without consideration to the variation that could take place in the feedstock quality. This depict the fact that that the volume of sample is not the best measure for feeding a reactor especially when there is possibility for a huge change in feed characteristics. The huge impact that an increase in VS concentration had on OLR could explain one of the reasons why perturbation occurs in existing full-scale AD as discussed in section 4.1.3.

5.1.2 Filamentous Index (FI)

The FI data were reported in figure 5-4 from where for the first 6 days and between day 10 and to 15, the FI in the feed sludge were fairly constant while it was varying greatly for the rest of the other days. This is related to the operating condition in the activated sludge treatment process. There was not much variation in the FI in the digesters except on day 7 and 11. This reflects that the reactors were good replicates of one another.

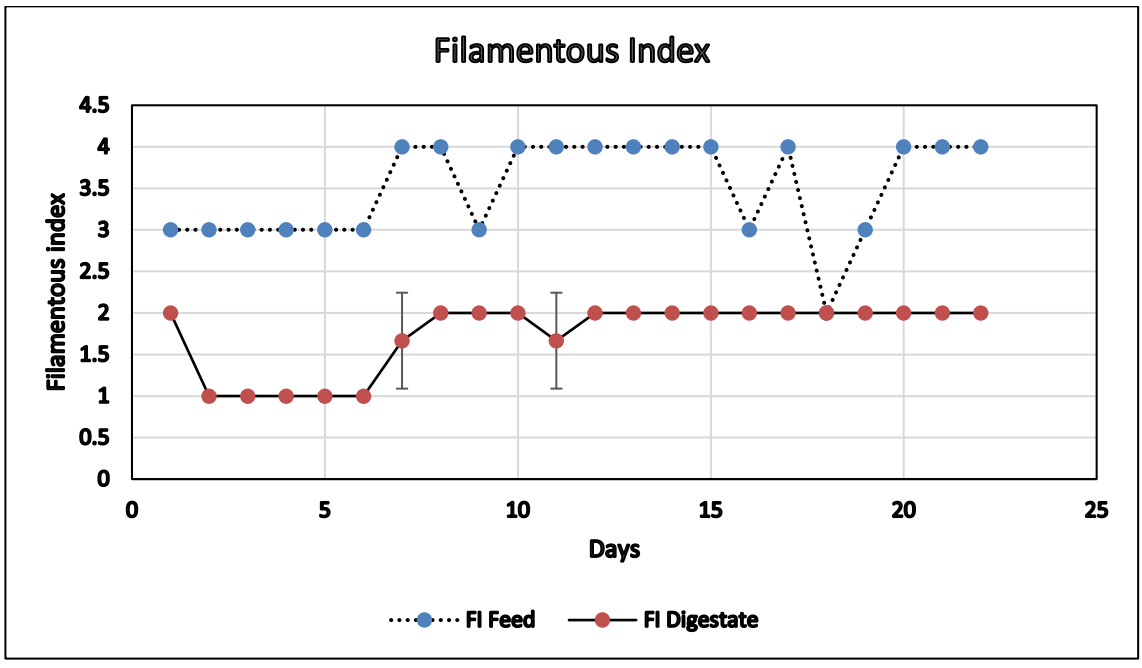


Figure 5-4: Variation in Filamentous Index

Referring to figure 5-4, the initial FI in the digesters were 2.0 as the inoculum was taken from full scale digester, but because of the higher efficiency in the reactor within the first days as shown in the high reduction in VS concentration in section 5.1.1, there was a reduction in the filamentous bacteria as the digester environment did not encourage it to thrive.

However, after the sixth day, there was an increase in the FI of the feed which had an impact on the digesters resulting in an increment in the FI of the digesters. The FI remained constant for the rest of the experiment irrespective of other changes in the FI of the feed. Thus, it could be inferred that the increase of FI in the reactor is more dependent on the feed characteristics rather than on the operating conditions although the later can escalate its proliferation.

The same was observed in a laboratory scale activated sludge systems when operated under regimes of continuous or intermittent feeding of substrate. Houtmeyers et al., (1980) found that continuously fed AS systems repeatedly resulted in the development of filamentous bacteria and bulking of the sludge while intermittently fed systems did form good settling sludge, without filamentous bacteria. (Houtmeyers, et al., 1980).

Figure 5-5 and 5-6 represent typical examples of gram stain for the feed and digestate sludge samples respectively.

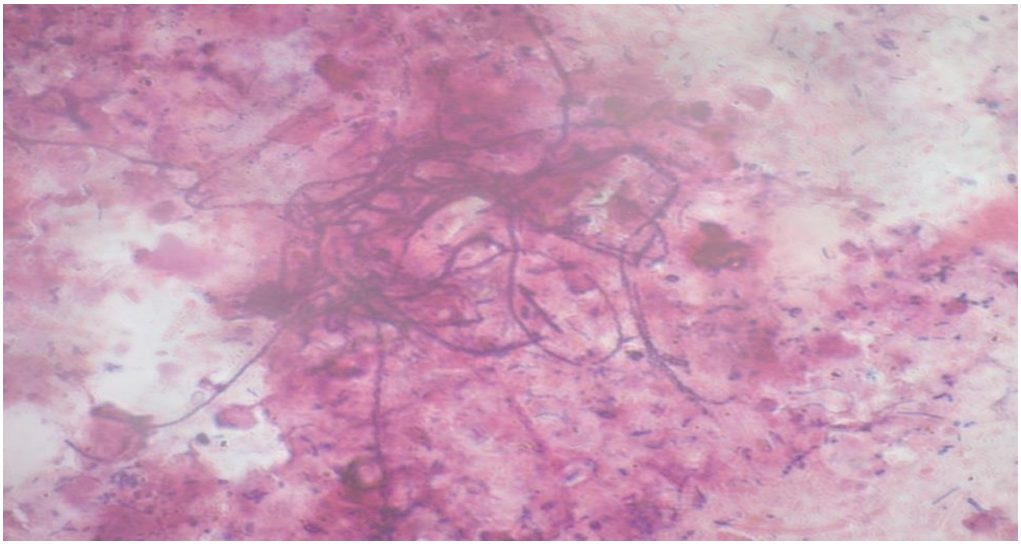


Figure 5-5: Typical example of the gram stain for feed sludge sample with FI of 3

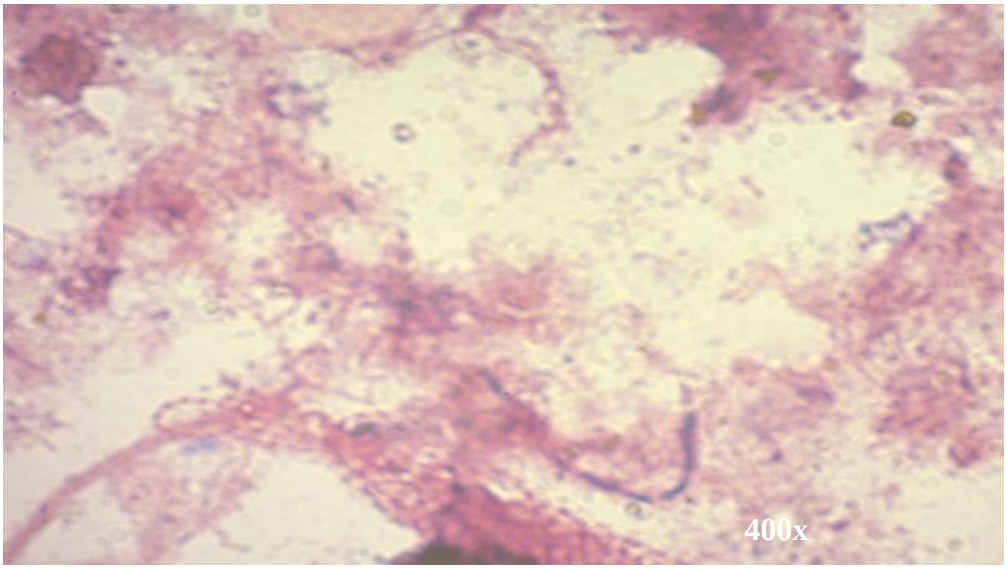


Figure 5-6: Typical example of the gram stain for digested sludge sample with FI of 1

The method applied was discussed in section 3.2.2.5; where blue and pink stain represent gram positive and gram-negative bacteria respectively.The pictures obtained were compared to the standard pictures for assessing the FI of gram stained sample as contained in Eikelboom (2000) base on which the FI were assigned.

A one-off analysis using Confocal Laser scanning microscopy (CLSM) was carried out for the feed and digester sample using fluorescent staining techniques which further support the gram stain test as shown in figure 5-7 and 5-8. Where red and green stain represent gram positive and gram-negative bacteria respectively.

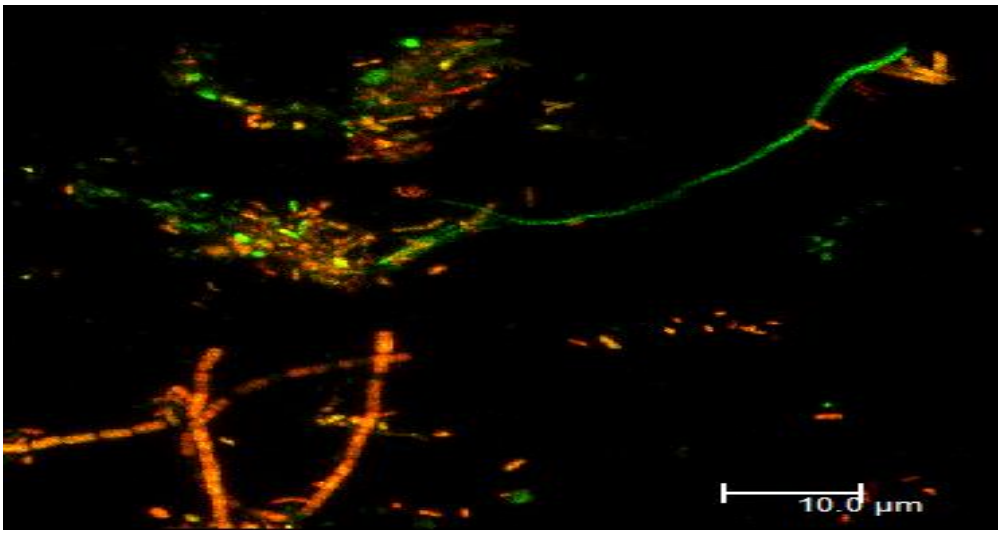


Figure 5-7: T Typical example of the CLSM_Gram Stain for feed sludge sample

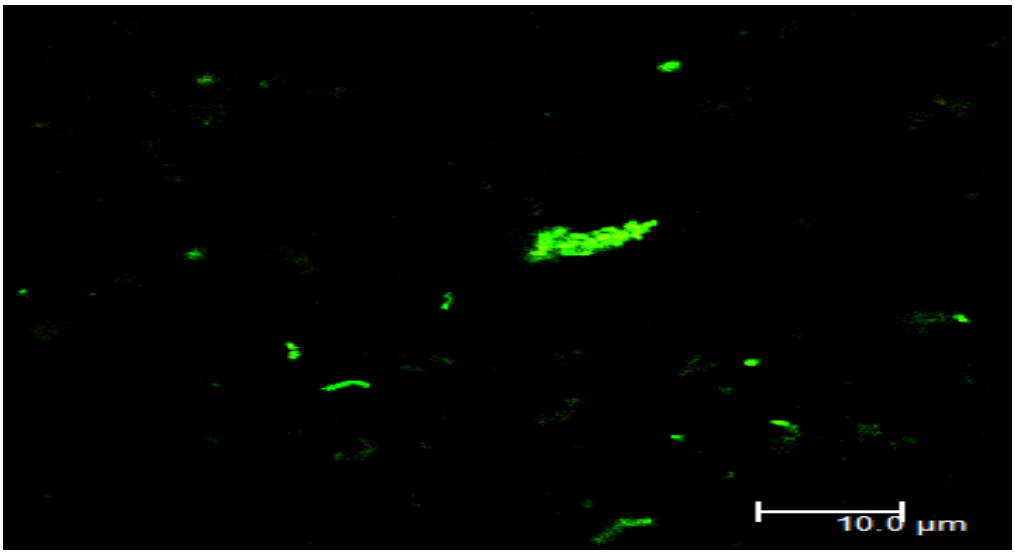


Figure 5-8: T Typical example of the CLSM_Gram Stain for feed sludge sample

5.1.3 Volatile fatty acid concentration (VFA)

Studies on influence of VFA on AD operation have been focused on digestate concentration without adequate consideration to the impact of VFA of the feed sludge on the digester operation as discussed in section 2.5.2. However, in section 2.8.1.4, Hill model highlighted the need for the impact of VFA concentration in the feed in evaluating the anaerobic digestion process, hence the determination of biodegradation and acid coefficient were the only two parameters required to characterise the digester feed for the different type of waste thereby generating a simplified Hill model.

Thus, in this study the VFA for the feed and digestate was carried out to monitor its impact on the digester efficiency. It could be seen from figure 5-9 that the VFA in the feed sludge was erratic peaking at day 13 and lowest at day 19. This relates with the VS which peaked at day 13. This could be explained on the basis that a massive increase in VS will be an indication that there will be an increase in VFA of the feed due to elevated rate of digestion in the feed sludge holding tank. In addition, there is potential that preceding the feeding of the digester, some facultative hydrolytic and acidogenic

bacteria could kick off digestion process prior to the digester been fed with sludge.

Furthermore, from section 5.1.1, it could be inferred that though there was a decrease in the volatile solid reduction, some of the biosolids was still degraded which led to further accumulation of VFA in the digester.

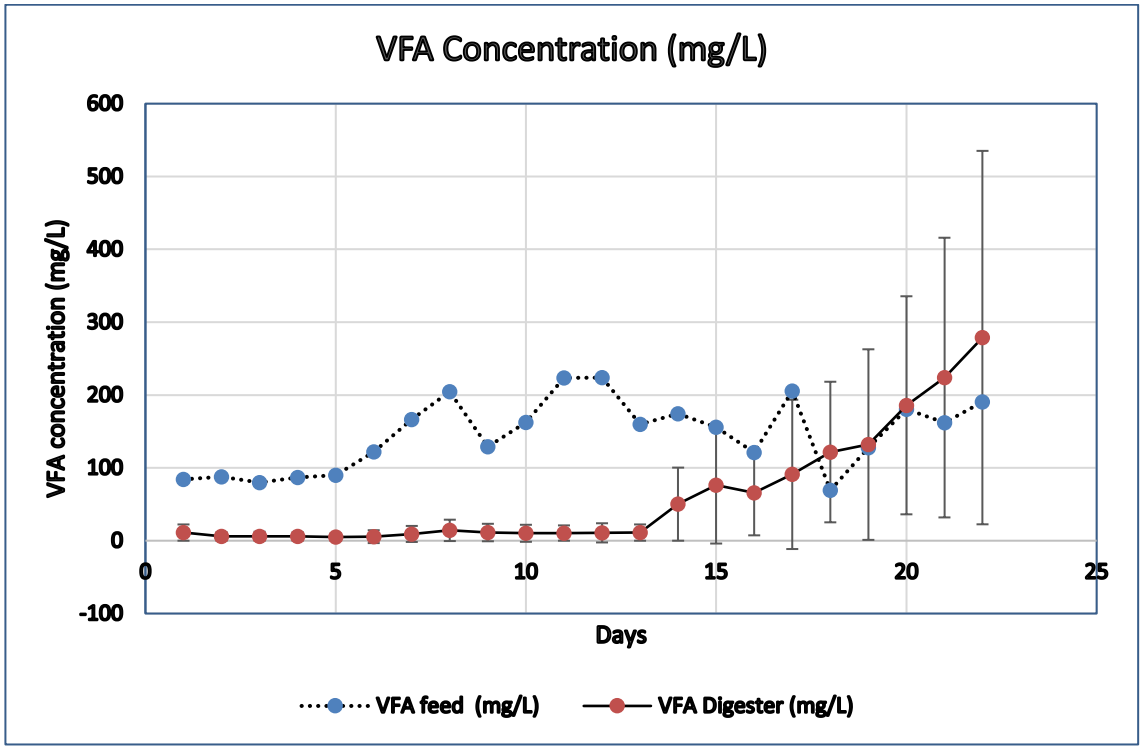


Figure 5-9: VFA variation concentration in feed and digester sludge

Though there were erratic variations in the VFA concentration of the feed sludge, it was observed that the VFA concentration in the digestate remained almost constant until day 13 when it commenced a continuous increment. This relates to the same day that the highest variation in VS concentration was observed as in section 5.1.1. It could be observed from figure 5-9 that as the values of VFA for the digester continued to increase, the variability in the performance of the digester was massive. Hence the reason most study has pointed out the deviation of the digester from the usual digester specific VFA concern as a good indicator of AD system unbalance.

In addition, it was observed that those points at which the VFA in the digestate coincided with the VFA in the feed and continued to increase. This

is as a result of accumulation of VFA in the digesters as the system operating conditions continue to deteriorate. It is also fascinating to notice that the increment in VFA commenced on the same day as the day there was a huge increment in the VS concentration thus reinforcing the fact that excess organic loading of reactors as one of the key causes of perturbation in anaerobic digestion process.

5.1.4 Biogas production rate (BPR) mm^3/hr

It could be seen from 5-10 that there was constantly a low variation in the amount of biogas produced by each digester. It means that the digesters functioning was fairly the same. Even within the foaming period in which there is a huge variation in VFA concentration there was continuous increment in the production of biogas.

This is showing clearly that at some point through the experiment, the biogas production rate was not proportional to volatile solid destruction. This could have happened at the points of foaming as most of the gas produced may not have escaped to the collection column as they are trapped in the foam.

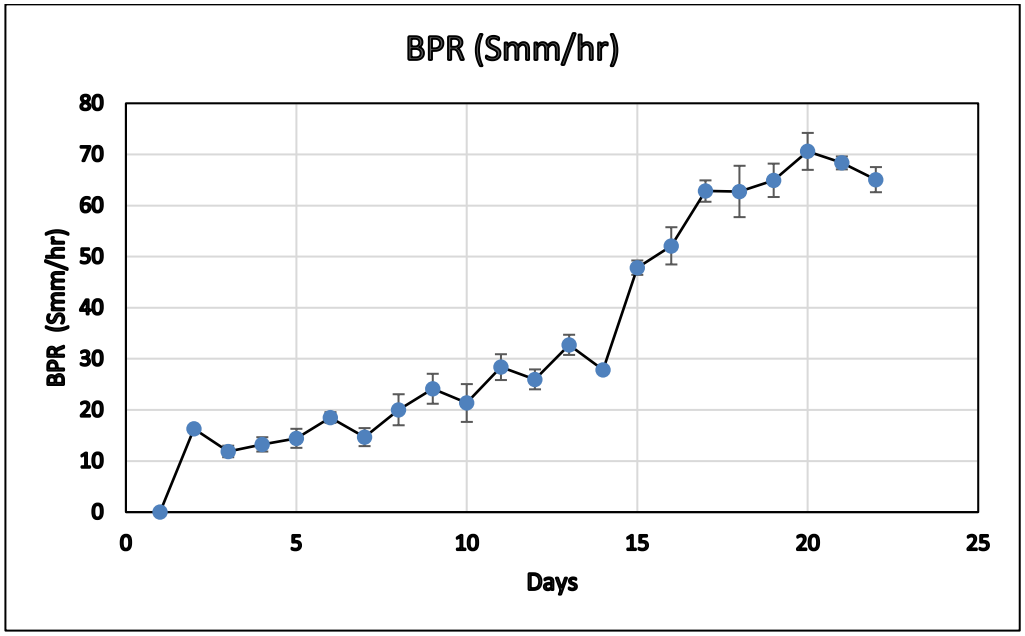


Figure 5-10: Biogas production rate PR

In addition, the high solubility of carbon dioxide allows it to dissolve in the digester once it could not escape easily thereby increasing the acidic content of the digester and making it more difficult for the digester to recover easily. As the digester experience challenge in to recovering to normal operation, the condition in the digester tend not to favour the methane forming bacteria resulting in the production of more carbon dioxide and hydrogen sulphide. Most of the hydrogen sulphide reacts with the moisture from vapour condensation forming sulphuric acid which flows back to the digester to further destabilise the digestion process.

5.1.5 Carbondioxide composition (%)

It could be observed from figure 5-11 that there is a huge drop in composition of carbon dioxide produced on day 13. Comparing this with figure section 5.1.4 BPR had a huge increment on the same day points to the fact that carbon dioxide production during optimal operation is low compared to when the system is going bad.

In addition, the occurrence of carbon dioxide production at different stages of the anaerobic digestion process such as; acidogenesis, acetogenesis and methanogenesis is reflected in the high disparity of values obtained in the early stage of the digestion process as shown in figure 5-11 at the. However, it could be observed at the later stage when the foaming occurrence is taken place, the carbon dioxide production was the same supporting the fact that more carbon dioxide is being produced at that point. These make it very difficult to use carbon dioxide as an effective marker to monitor the efficiency of the digestion process.

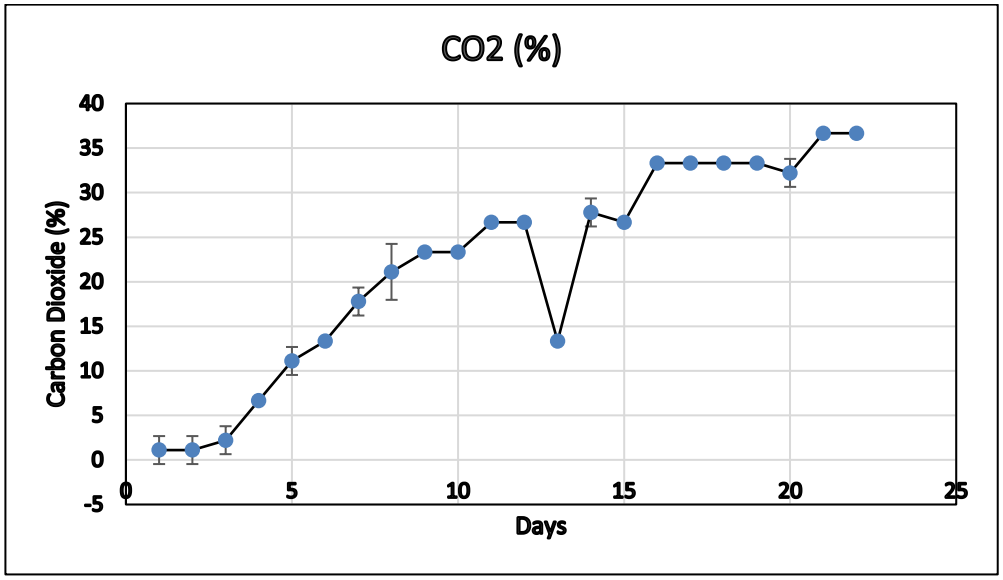


Figure 5-11: Carbon dioxide composition

5.1.6 pH

Based on figure 5-12, it was observed that the pH was not varying much for the different digesters which means that the digesters were operating within the same range of pH. Nevertheless, it is worth noting that slight changes in pH could have an adverse effect on the digestion process. This due to the fact that pH values represent a cumulative change in the digester condition hence pH depression does not occur until alkalinity is depleted. Consequently, pH is not a very efficient variable to respond quickly to digester disturbance. However, it is one of the easiest variables to determine thus consistent monitoring of the pH could be a useful guide in identifying an anomaly in anaerobic digestion process.

Since 75% of the pH values were greater than 6.85 for the digesters. This shows that the digesters were operating at an optimal pH condition most of the time during the experiment as digesters can operate at pH ranges from 6.0 to 8.0, with the desired operating range being 6.8 to 7.2 (Eddy/AECOM, 2014). However, if the pH falls below 6.0, un-ionized volatile acids become toxic to methane-forming microorganisms. At pH values above 8.0, un-ionized aqueous ammonia (dissolved ammonia) becomes toxic to methane-forming microorganisms. The maximum pH of 7.15 was is in accordance with the stipulated value of 7.2 for optimal digester operation.

As reported in figure 5-12, the pH of the reactors gradually increased to a maximum value of 7.15 on day 9 after which it varied for some time before it commenced a gradual decline as the operation condition deterioriates. Since the pH of the digester contents is controlled by the acids and alkaline concentrations in the digesters, it is apparent that at the early stage of digestion when there is adequate balance in the acid/alkaline ratio, the digester pH were farily the same but when the sysytem changes especially from day 13 when there was a drastic increase in OLR and VS concentration, the pH values for each reactor were dissimilar.

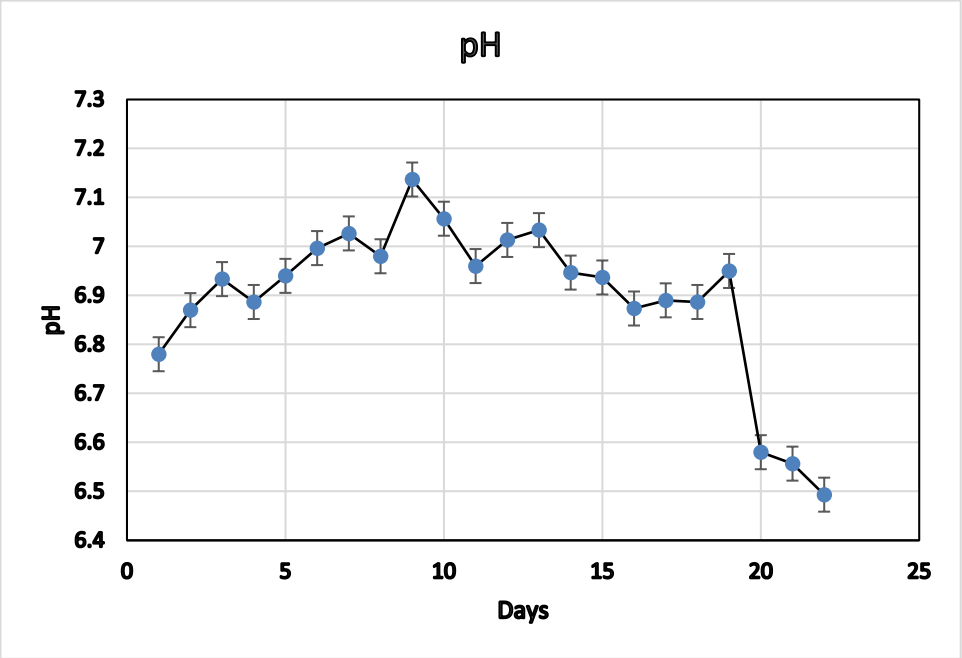


Figure 5-12: pH measurements

These values represent the foaming period in the digesters when the pH dropped low from the usual operating pH by 5.4%. However, there was not a signiificant drop in the pH until day 20 when the system was already getting sour thus indicating that pH is not a proactive measure to monitor AD pertubation .

5.1.7 Foaming

5.1.7.1 Additional digester volume

The additional digester volume was determined as explained in section 3.2.3.2. As reported in figure 5-13 the additional digester volume ranged from a minimum of 0% to a maximum of 40% and there was no variation in the additional digester volume covered by foam for the different digesters.

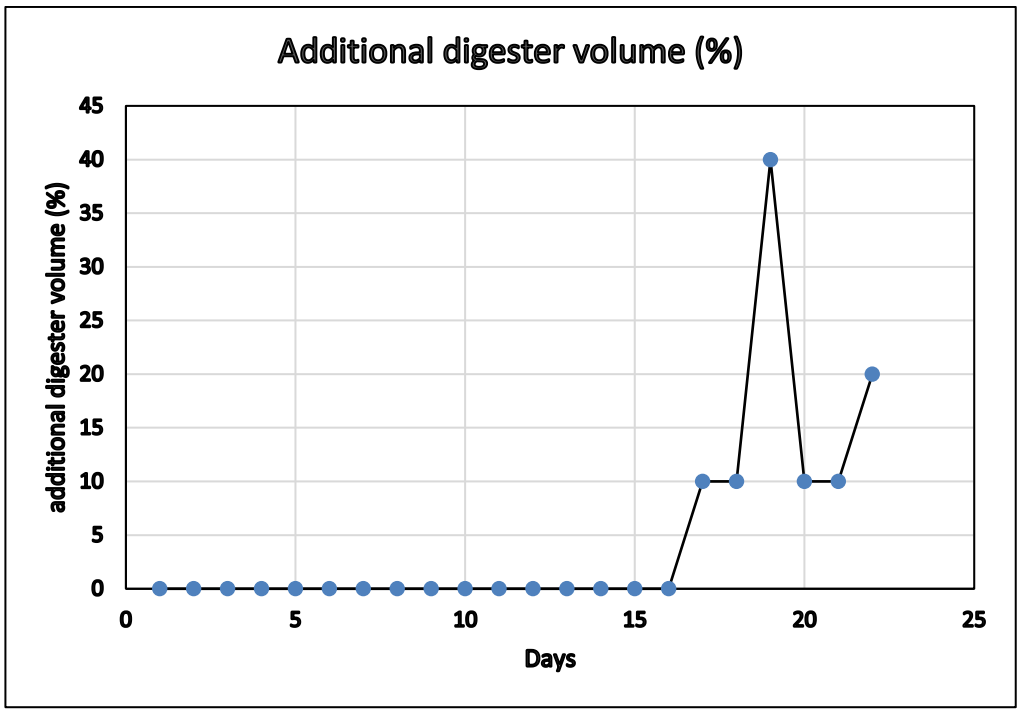


Figure 5-13: Variation in Additional digester volume occupied by foam in digesters D, E and F

Following the 5.4% drop in pH in figure 5-12 there was a spontaneous increase in the foaming in the digester resulting to a 75% increment in the additional digester volume from the existing volume occupied by foam in the digesters which further confirm that it is at the point that the digester is getting sour that the pH reflects a change in the AD operation.

However, it could be observed from previous discussions that VFA increase started from day 13, on the same day there was a drop in the carbon dioxide composition in the biogas produced and an increment in VS concentration resulting in an increase in OLR and biogas production. Thus, this signifies that that VFA is the quickest indicator of system perturbation in the digester.

Hence a combination of knowledge of VS and VFA in the feed could be very useful in avoiding AD nuisance foaming.

5.1.7.2 Maximum volume of foam

Intrinsic nature of anaerobic digester content of produced biogas and surfactant (bio surfactants and VFA) warrants that foaming occurs more often than it is noticed as it may not result to digester overflow. In most cases the digester operating condition could be restored to normalcy without causing any nuisance.

However, when the foam overflows from the digester, then, it is recorded that there is foaming occurrences by full scale anaerobic digester operators. Maximum volume of foam as recorded in this experiment reflect these occasions of foaming without digester overflow as well as those occasions when possible digester overflow must have been witnessed. As noted earlier and confirmed from table 5- 8, it could be seen that the digesters did not experience foaming most of the time during the experiment as 75% of the values for maximum volume of foam was recorded as zero. Digesters E and F with the same mean and standard deviation seems to be operating almost at the same condition.

In addition, figure 5-14 and 5-15 show that the foaming occurrence in the reactors were quite similar except for day 13 that there was a variation in the digester foaming which based on the preceding discussion in section 5.1.7.1 is related to the onset the AD perturbation.

The data for the maximum volume of foam recorded in the digesters were converted as percentage of digester volume and plotted in figure 5-15. This was essential as it formed a basis for easy comparison of data from different AD sizes as well as allows for efficient adaptation of the foam predictive model for different digester size.

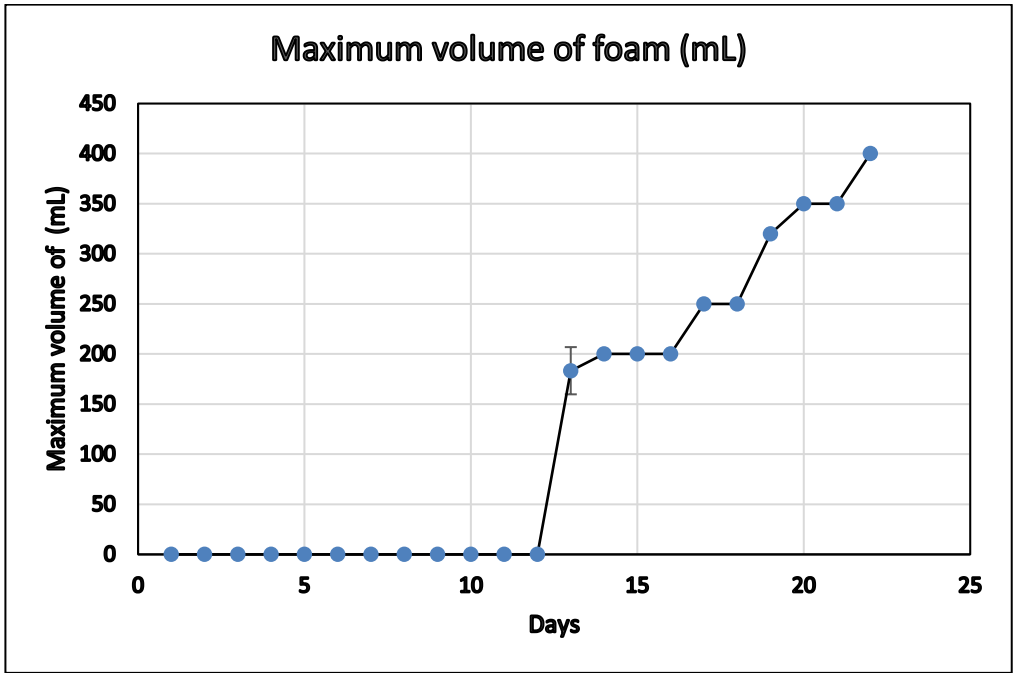


Figure 5-14: Maximum volume of foam (mL)

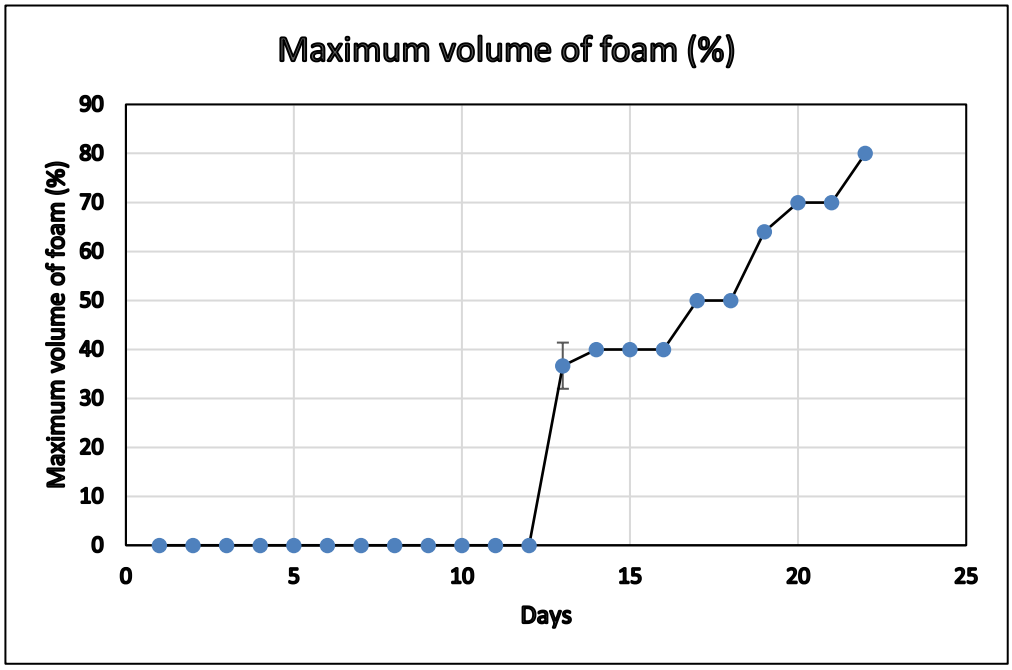


Figure 5-15: Maximum volume of foam (%)

5.2 Deductions from the experiment

It could be observed from discussion in section 5.1 that some of the variables relate more with the other. When VSR and pH were decreasing, there were increase in OLR and VFA. Thus, increase in organic loading rate resulted to a decrease in the percentage of volatile solid reduction achieved through the

anaerobic digestion which could be related to a drop in the operating conditions in the digester as depicted by a drop in the digester pH resulting from a rise in VFA.

The FI and VFA of the feed sludge were exhibiting some relationship which could be explained from the point of view that facultative hydrolytic and acidogenic bacteria present in the feed sample will start acting on the feed leading to a progression in the digestion process prior to feeding to the digester. Hence with a higher bacteria presence in the feed, irrespective of whether it is filamentous or not may lead to a higher VFA in the feed. This is the main reason why the inclusion of the VFA concentration in the feed and the OLR is necessary to ensure that the digesters could function optimally.

Volume of feed, organic loading rate, FI of feed, BPR, FI of digestate and VFA in the feed digestate exhibited the same trend with percentage of additional digester foam and maximum volume of foam showing that there exist some directly proportional relationship between these factors and foaming occurrence. While the relationship between foaming occurrence and % of VS reduction/ pH is inversely proportional. The % of CO2 in the biogas composition seems to be the only variable solely on its own as its pattern did not relate with any other variable. This further illustrates that carbon dioxide production increases during system perturbation and adding to its high solubility, it becomes difficult to adequately utilise it a measure to efficiently monitor AD operation.

5.3 Normality test

With the identified relationship as discussed in section 5.2, it was then necessary to assess the statistical significance of the identified relationships. However, to ensure that the most suitable method of assessment was applied, the normality of the distributions was determined.

This was carried out as discussed in section 2.8.1.3.1 using equation 2-8 and 2-9 for determining the absolute Z-score for skewness and Kurtosis and

summarised in table 5-1. The standard error for skewness and kurtosis were determined as 0.52 and 1.04 respectively.

Table 5-1: Assessment for the normality of data distribution

Variables	Mean	STD	Kurtosis	Skew	Absolute Z-score kurtosis (Kurtosis /Std error)	Absolute Z-score skewness (Skew/Std error)	Alpha level
VFA (mg/l)	145.79	48.88	-1.19	-0.02	1.14	0.04	0.05
OLR (kg/m^{3}d)	3.02	2.23	1.36	4.98	1.31	0.95	0.05
FI Feed	3.50	0.60	-0.31	-0.74	1.92	1.42	0.05
VSR (%)	60.54	7.36	2.2	-1.29	2.12	2.48	0.05
BPR (%)	34.73	22.44	--1.35	0.43	1.30	0.83	0.05
pH	6.90	0.16	1.48	-1.35	1.42	2.60	0.05
Additional Dig. Volume (%)	4.55	9.63	8.69	2.79	8.36	5.36	0.05
MVF (%)	24.58	29.40	-1.29	0.59	1.24	1.14	0.05
CO_2 (%)	21.87	11.83	-0.96	-0.55	0.92	1.06	0.05
VFA (Digestate) mg/l	61.02	80.14	1.65	1.56	1.58	3.01	0.05
FI (Digetstate)	1.74	0.42	-0.35	-1.24	0.34	2.38	0.05

The data summarised in table 5-1 show that data for most of the variables monitored during the experiment were normally distributed. Hence, correlation analysis was selected to be best suited for estimating the statistical significance of the relationship existing between the variables.

5.4 Correlation analysis

The result of the correlation analysis is shown on table 5-2. Comparing the statistical significance of the various variables in relation to foaming occurrence formed the basis for selecting variables that was best suited for developing predictive model of foaming in AD.

Table 5-2: Correlational analysis of the variables

	Volume of feed sludge (mL)	OLR ($kgVS/m^3$)	VFA feed (mg/L)	VFA digestate (mg/L)	FI feed	FI digestate	VS feed (%)	VS digestate (%)	VSR (%)	BPR (Smm^3/hr)	pH	ADV (%)	MVF (%)
OLR ($kgVS/m^3$)	0.990												
	0.000												
VFA feed (mg/L)	0.422	0.443											
	0.050	0.039											
VFA digestate (mg/L)	0.880	0.861	0.311										
	0.000	0.000	0.158										
FI feed	0.394	0.403	0.878	0.303									
	0.070	0.063	0.000	0.171									
FI digestate	0.776	0.773	0.632	0.726	0.525								
	0.000	0.000	0.002	0.000	0.012								
VS feed (%)	0.608	0.577	0.252	0.540	0.324	0.574							
	0.003	0.005	0.257	0.010	0.141	0.005							
VS digestate (%)	0.922	0.902	0.360	0.816	0.269	0.679	0.525						
	0.000	0.000	0.100	0.000	0.226	0.001	0.012						
VSR (%)	-0.555	-0.539	-0.059	-0.577	-0.043	-0.158	0.120	-0.668					
	0.007	0.010	0.793	0.005	0.849	0.483	0.594	0.001					
BPR (Smm^3/hr)	0.966	0.949	0.429	0.869	0.370	0.768	0.593	0.933	-0.566				
	0.000	0.000	0.047	0.000	0.090	0.000	0.004	0.004	0.006				
pH	-0.263	-0.228	0.247	-0.426	0.161	0.138	-0.041	-0.302	0.677	-0.294			
	0.237	0.307	0.267	0.048	0.475	0.539	0.855	0.172	0.001	0.184			
ADV (%)	0.765	0.763	0.141	0.770	0.055	0.407	0.368	0.765	-0.666	0.762	0.495		
	0.000	0.000	0.532	0.000	0.809	0.060	0.092	0.000	0.001	0.000	0.019		
MVF (%)	0.912	0.897	0.179	0.894	0.224	0.580	0.497	0.875	-0.730	0.897	-0.541	0.842	
	0.000	0.000	0.427	0.000	0.317	0.005	0.019	0.000	0.000	0.000	0.009	0.000	
CO_2 (%)	0.936	0.909	0.446	0.853	0.328	0.777	0.608	0.919	-0.532	0.911	-0.265	0.733	0.831
	0.000	0.000	0.038	0.000	0.136	0.000	0.003	0.000	0.011	0.000	0.234	0.000	0.000

As shown from the result reported in table 5-2, the top and bottom column for each box represent the correlation and p value respectively. Based on the result of the correlation analyses, all the relationships with p-value less than 0.05 with a high correlation coefficient greater than 5.0 were selected as highly correlated and exhibiting statistically significant relationship respectively.

The relationship between organic loading rate and volume of feed with S=0.990 and p= 0.000 is expected as organic loading rate was calculated based on the volume of feed, however, it was used in this analysis to identify which one of them relates more with foaming occurrence. It will also be identified that additional digester volume occupied by foam (%) and

maximum volume of foam (%) observed in the digester was also used in this analysis to ensure that the best variable that defines the foaming episode will be selected for the predictive model.

There was also a good correlation between additional digester volume occupied by foam (%) and maximum volume foam with an S= 0.842 and p = 0.000. There exists an inverse relationship between the VSR reduction with the volume of feed and organic loading rate as identified in the scatter plot with S of -0.555 and -0.539 respectively. This shows that as the organic loading rate is increased, the digester finds it difficult to cope with the changes in the loading and there is a reduction in the breakdown process which reflects a decrease in the VSR of the digester.

The result also shows that biogas production rate is directly related to organic loading rate but inversely related to VSR. An increase in organic loading rate increases the available organic matter that is needed to be broken down to generate biogas thus should increase the biogas production. It should be expected that a decrease in the VS reduction should result to a decrease in the biogas production, however, this is in the contrary and could be related to the fact that even though there is a decrease in the VS reduction, the anaerobic digestion process allows for the production of other gasses such as hydrogens sulphide and carbon dioxide while lowering the methane content of the biogas.

The pH was not significantly related to any other variable except VSR reduction with an S of 0.677 and p of 0.000. This is quite in line with the fact that the digesters perform better at the higher optimal pH range of 6.8 to 7.2, hence the reduction in the % of volatile solids experienced by the digesters. However, its inability to relate with other variables highlights that it is not a very efficient marker for an early detection of foaming in anaerobic digester.

Maximum volume of foam in the digester displayed the same pattern of relationship with other variables as the additional digester volume occupied by foam, though with a higher correlation coefficient. The relationship between the foaming occurrence variable and organic loading rate concur with the notion that an increase in organic load to the digesters over time will

have a detrimental effect on the digestion process which could lead to foaming occurrence. It could also be seen that an increase in foaming result to a decrease in pH and VSR. This shows that lower pH values are not suitable for digester operation and could reduce the digestion and result to foaming problem.

Carbon dioxide composition in the biogas seems to be highly correlated with most of the variables. However, due to the conditions that promote high production of carbon dioxide and its high solubility, it does not represent a very useful variable for a predictive model. VFA concentration in the digestate exhibited a very high statistically significant correlation with most of the variables used for this analysis indicating that it is very useful in assessing the digester performance. An increase in the digestate FI relate with an increase in the organic loading rate, BPR, carbon dioxide composition and VFA concentration in the digestate at a statistically significant level.

5.5 Conclusion

In the preceding discussions, the operating features of AD have been explored and attempt made at establishing the relationship that exist between these variables during foam formation in order to understand how they influence AD foaming.

It was observed that OLR, volatile solid concentration in the digestate and biogas production rate were the variables that exhibited the most statistically significant relationship with AD foaming. However, OLR was the only input variable to the digester while the VS in the digestate and VFA are output variables from the digester thus will not suffice as variable for developing a proactive prediction of foaming in AD.

On the other hand, the VFA in the feed could be a very useful factor that though did not have a statistically significant relationship with foaming has a significant relationship with FI which reflect the key microbial contribution to foaming. Furthermore, it has been shown from the discussions that the increase in FI increases VFA in the feed due to the potential activities of microorganism prior to entering the digester and which could have an overall

instantaneous increase in the AD system dynamics resulting to increase in VFA concentration in the digester that will lead to system perturbation.

Hence, following on the extensive knowledge gained from the outcome of this experimental study, ‘omics’ analysis was carried out to further verify the result obtained and ensure that the variables selected for the predictive model will be an excellent cursor of foam formation in AD.

Chapter 6 : 'Omics' analysis

'Omics' methods are high-throughput and data-driven molecular techniques that are carried out for the whole microbial population without the need for any separate culturing of specific strains. Consequently, speedy advances in these molecular technologies paved way for the combination of 'omics' approaches such as metagenomics, metaproteomics, metatranscriptomics and metabolomics to provide invaluable information on the genetic, functional and metabolic activities of microbial communities. However, each technique is aimed at diverse constituents of the microbial community. Thus, applying knowledge gained from full scale AD study and laboratory experiment executed, metagenomics and metabolomics analysis were selected as the most suitable 'omics' analysis to decipher the contribution of microorganism and metabolite to foaming in AD and identifying the best indicator for predicting foam formation in AD.

The need to carry out metagenomics analysis was essential to determining the collection of all genomes present in the biological samples in order to infer their contribution on the foaming occurrence in AD. On the other hand, Metabolomics analysis were used to establish the intermediate and end products of cellular metabolism with the ability to identify the most accurate occurrences of basic metabolic. This is as a result of the fact that metabolic fluxes are not regulated by gene expression or posttranslational modification (Hettich et al., 2013) as obtainable in metatranscriptomes and metaproteomes.

6.1 DNA Quantificiation

Following the procedure set out in section 3.2.5.1, DNA extractions were carried out and quantified. Table 6-1; highlight the concentration of DNA extracted for each sample which was within the range necessary for the genome sequencing.

Table 6-1: Typical DNA Extraction quantification using Nanodrop 2000c for the metagenomics sample

Sample ID	Nucleic Acid Conc. (ng/µl)	A260	A280	260/280	260/230
Innoculum	8.7	0.175	0.088	2	0.46
Feed	64	1.279	0.604	2.12	1.21
Non-foaming before feed	8.3	0.166	0.076	2.2	0.64
Non-foaming	13	0.26	0.129	2.01	0.73
foaming before feed	24.6	0.492	0.246	2	1.46
foaming	33.5	0.671	0.326	2.06	1.09

6.2 Metagenomic analysis

To assess the biological variation and reproducibility of the samples, PCA was performed and the principal component scores plotted as shown in figure 6-1 for the set of experimental biological classes. On average, there were less variability between the samples thus are reproducible except for the sample for non-foaming digester before feed. This could be attributed to the elevated level of activity going on in the digester thereby varying its microbial population. This highlights the complexity in modelling microbial presence in AD especially for a very active digester comprising various microorganisms operating and proliferating at varying rates.

Another observation is that the PCA plot showed that the microbial population in the non-foaming digester increased after the digesters were fed while the microbial population for the foaming digester reduced after the feed. This further illustrate that there is existence of adequate operating condition for the microorganism to feed and proliferate in the non-foaming digester. Again, the feed sample had the lowest composition of microorganism compared to the rest of the sample

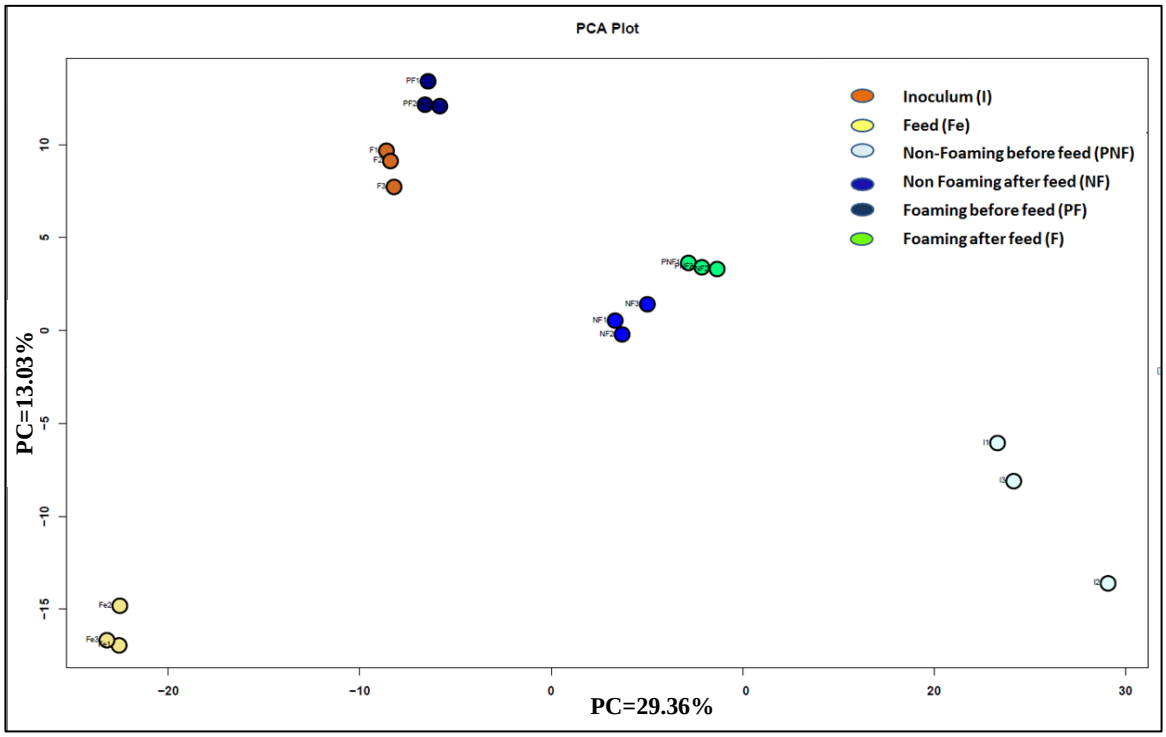


Figure 6-1: Plot of the first two principal components calculated for experimental biological samples used in the Metagenomics analysis

Figure 6-2 represents the Chao-1 diversity plot. As the name implies, it is focused on observing the diversity of the microbial population thus clarifying the difference between high compositions of one particular microorganism to the presence of diverse microorganism in a particular biological sample. It could be observed from the plot that the non-foaming after feed sample had the highest value showing that there are lot of single read of operational taxonomic unit (OTU) present in the sample while the feed sample had the lowest values showing the huge abundance of the different bacteria in the sample thus having a higher read for each OTU that is found in such sample and less diversity. Supporting this notion is the high relative abundance of Bacteriodietes (42.9%) and proteobacteria (48.5%) present in the feed sample compared to other samples.

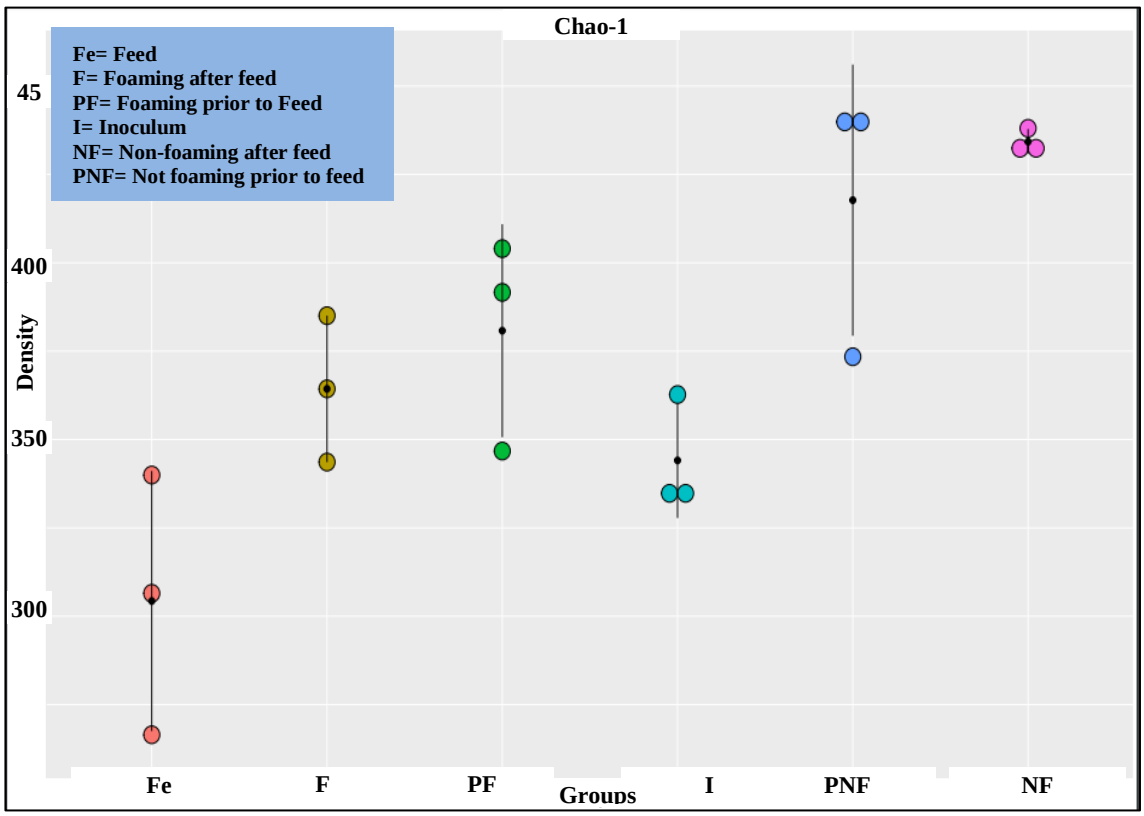


Figure 6-2: Chao-1 diversity plot

On the other hand, the Shannon index was plotted to effectively offer a clearer explanation concerning species richness and evenness in a given ecosystem. As the index implies, increases in the number of species or their evenness results in a better biological diversity, it could be viewed from figure 6-3 that the non-foaming sample after feed had the highest Shannon Index.

These highlight the fact that the non-foaming sample has a more suitable diversity of microbial population necessary for the AD process. However, the feed sample had the least Shannon index which support the findings that there is absence of methanogens in the feed sample (figure 6-3) because, methanogens are strictly anaerobic thus the feed sample coming from the activated sludge process which is basically aerobic is devoid of any methanogen.

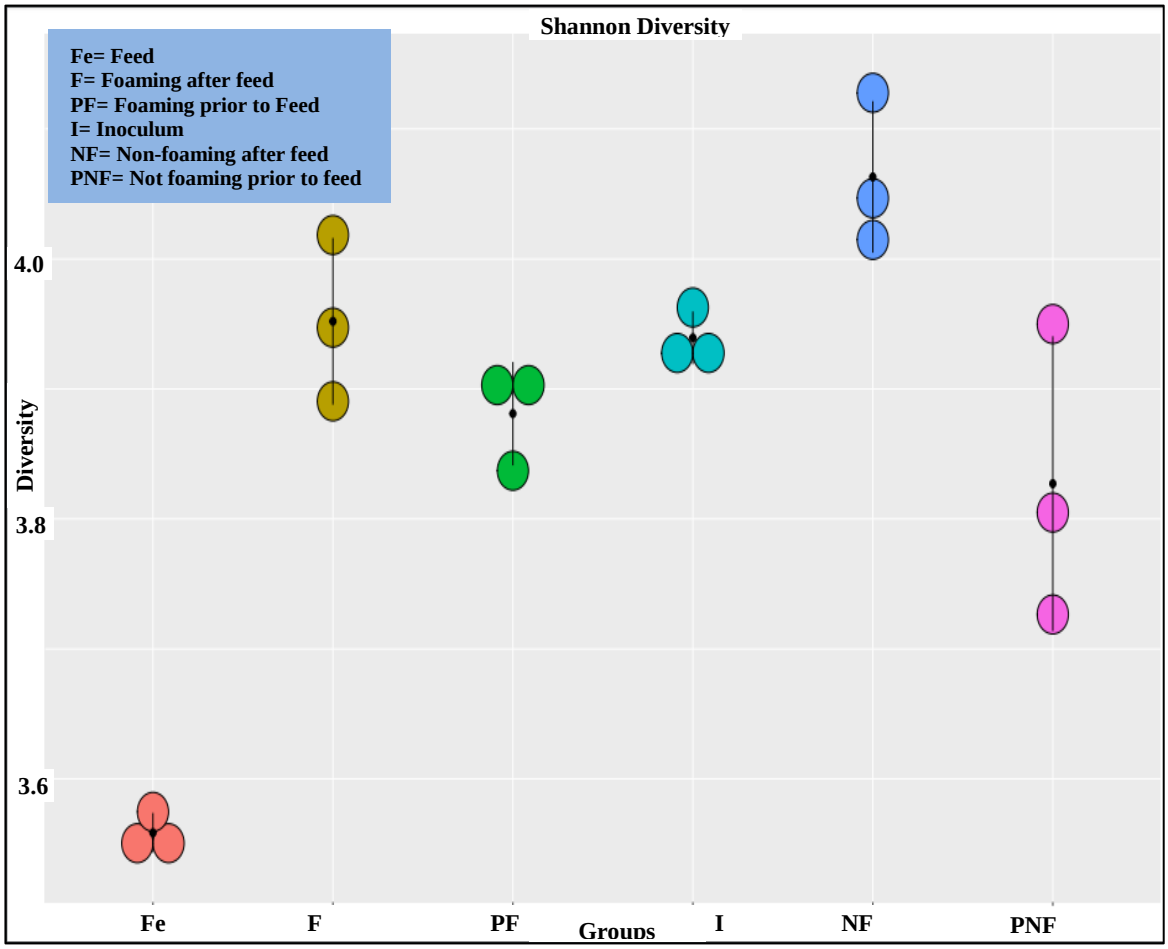


Figure 6-3: Shannon diversity plot

Furthermore, the result of the metagenomics analysis was presented as relative abundance of other microorganism in the sample with taxonomical classification from Kingdom to genus. In the following discussion, the relative abundance for each sect (three samples) of experimental class were averaged. It was observed that it was only the Kingdom Archaea and Bacteria that were present in the samples.

The Archaea kingdom was made up of only phylum Euryarchaeota which had an average relative abundance of 0.2%. For the Euryarchaeota, the relative abundance was higher in the inoculum (0.5%) and non-foaming digesters (0.4%) than in the foaming digesters (0.2%) while there was absence of it in the feed samples as shown in figure 6-4. Since Euryarchaeota are majorly methanogens which are mainly anaerobic, this support the fact that methanogen proliferates and survive less in aerobic as well adverse operating conditions.

As shown in figure 6-4, the relative abundance of the Archaea was more in the non-foaming digesters than in the foaming digesters. This further illustrates that methanogens are very susceptible to unfavourable conditions that exist in the AD during foaming occurrence such as accumulation of VFA.

R.M.W Ferguson et al. (2016) established links between the microbial composition and VFAs profile as analysis of the microbial communities indicate that changes in OLR induced changes in the microbial community structure and abundance.

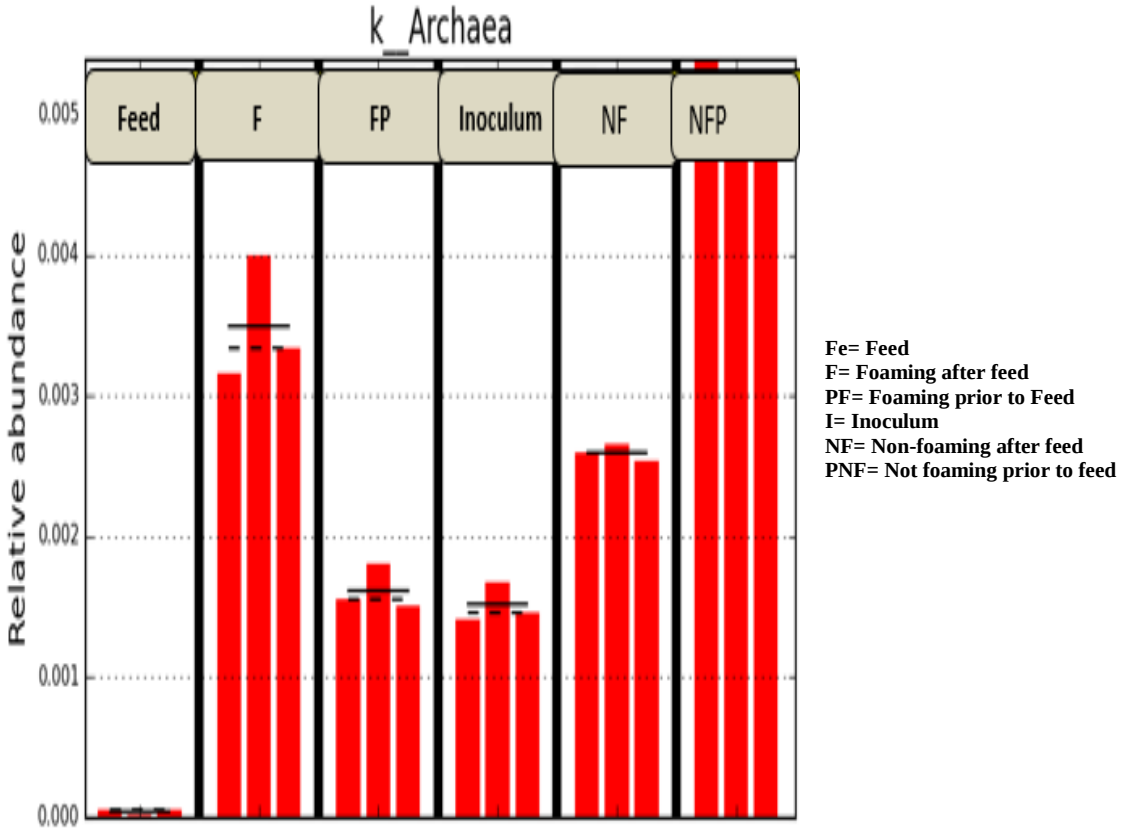


Figure 6-4: Relative abundance of Archaea in the different biological samples

In figure 6-5, the various biological samples had a high abundance of bacteria supporting the fact that all the biochemical stages in AD process (hydrolysis, acidogenesis and acetogenesis) except methanogenesis is controlled by the bacteria making up the microbial population. This affirms the need to ensure that there is an adequate balance in the ratio of bacteria and archae in the digesters for effective AD operation. As can be seen from figure 6-5, the relative abundance for all the samples were the same showing that bacteria

are more resistant at various operating conditions. Thus for efficient AD operation, the focus will be on controlling AD process is to make sure that the system operating condition permit the proliferation of methanogenic bacteria.

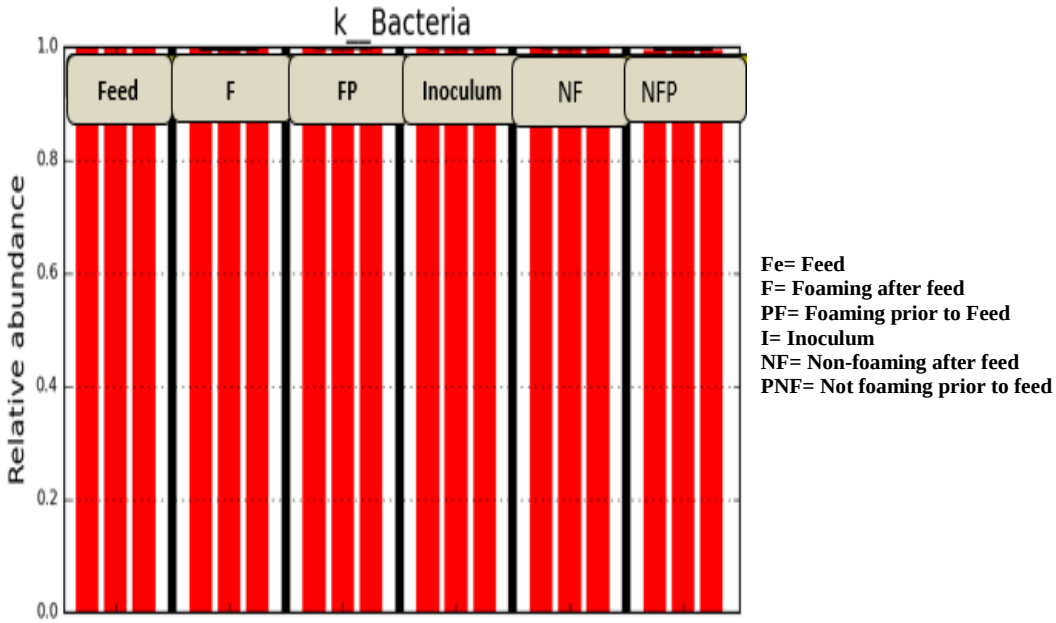


Figure 6-5: Relative abundance of Bacteria in the biological samples

In the bacteria kingdom, Bacteroidetes were the most abundant (30%), followed by Proteobacteria (29%) and Firmicutes (10% while the rest were below 10%.

Bacteroidetes are very diverse and have colonized virtually all types of habitats on Earth. They are among the major members of the microbiota of animals, especially in the gastrointestinal tract, can act as pathogens and are frequently found in soils, oceans and freshwater. In these contrasting ecological niches, bacteroidetes are increasingly regarded as specialists for the degradation of high molecular weight organic matter, i.e., proteins and carbohydrates. A shown in figure 6-6, the relative abundance of bacteroidietes were highest in the feed samples (42.9%) followed by the foaming samples (30.1%) and then the non-foaming samples (27.8%) which was almost the same as the inoculum (26.5%).

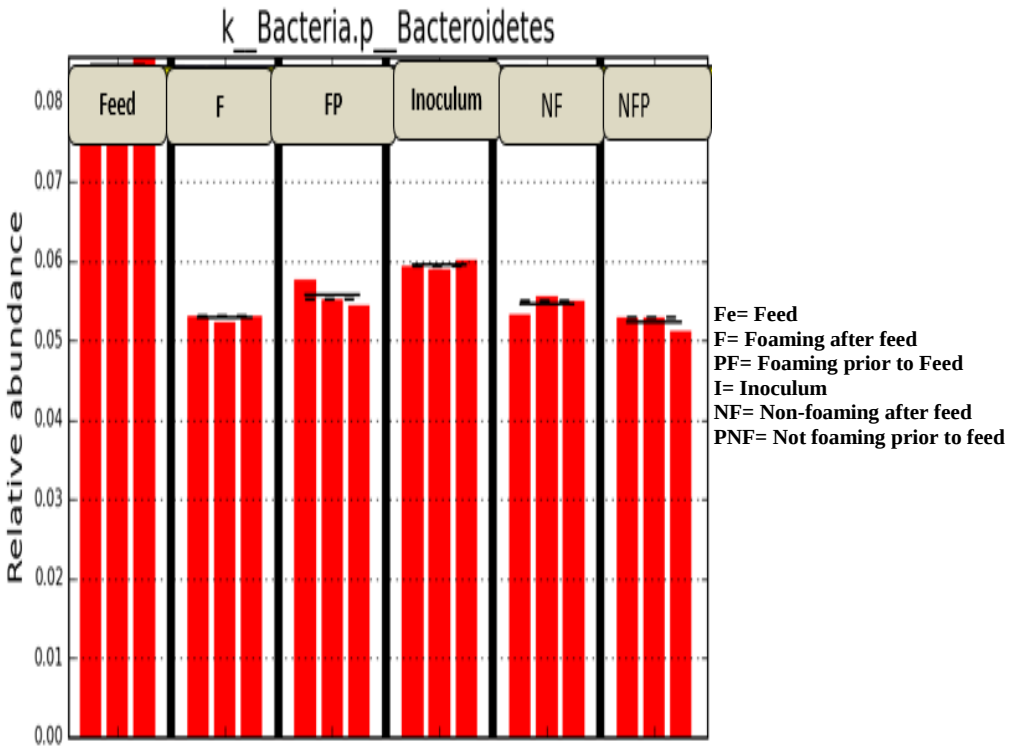


Figure 6-6: Typical Relative abundance of Bacteroidetes in the different biological sample

Proteobacteria is a diverse phylum of gram-negative bacteria that are mostly pathogenic. It includes many intestinal bacteria (e.g., Escherichia coli, Salmonella), the nitrogen-fixing bacteria, and the anaerobic purple bacteria. The relative abundance of proteobacteria were highest in the feed samples (48.5%) followed by the foaming samples (29.8%) then the non-foaming samples (28.5%) and finally the inoculum (17.8%). It could be observed from figure 6-7 that the anaerobic digester process is very efficient in the destruction of pathogenic microorganism as the relative abundance of the proteobacteria though very high in the feed sample was drastically reduced both in the foaming and non-foaming digesters.

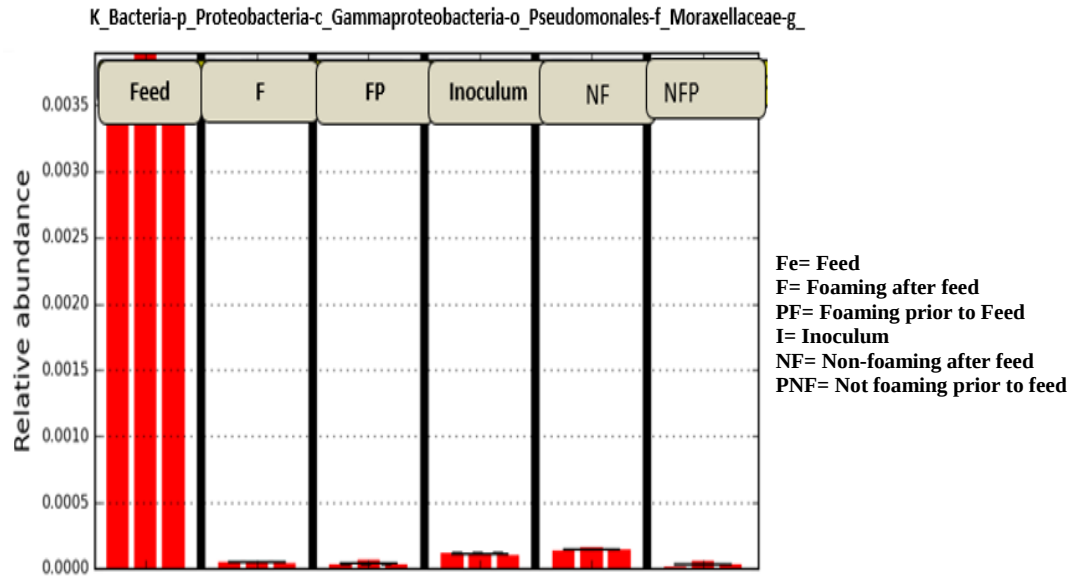


Figure 6-7: Relative abundance of proteobacteria in the different biological samples

The Firmicutes have strong gram-positive cell wall. A few, however, such as Megasphaera, Pectinatus, Selenomonas and Zymophilus, have a porous pseudo-outer-membrane that causes them to stain Gram-negative. They belong to a core group of related forms called the low-G+C group with round cells, called cocci (singular coccus), or rod-like forms (bacillus).

Many Firmicutes produce endospores, which are resistant to desiccation and can survive extreme conditions. The non-spore forming ones are non-respiring, can survive in very low pH conditions, rod shaped bacilli or cocci commonly used in the food industry, probiotics and biotechnology. They are found in various environments, and the group includes some notable pathogens. Those in one family, the heliobacteria, produce energy through photosynthesis. Contrary to existing trend, as shown in figure 6-8, the feed had the least abundance of firmicutes (4.3%) followed by non-foaming sample (7.2%) then inoculum (8.7%) and finally foaming sample had the highest abundance of firmicutes (16.6%).

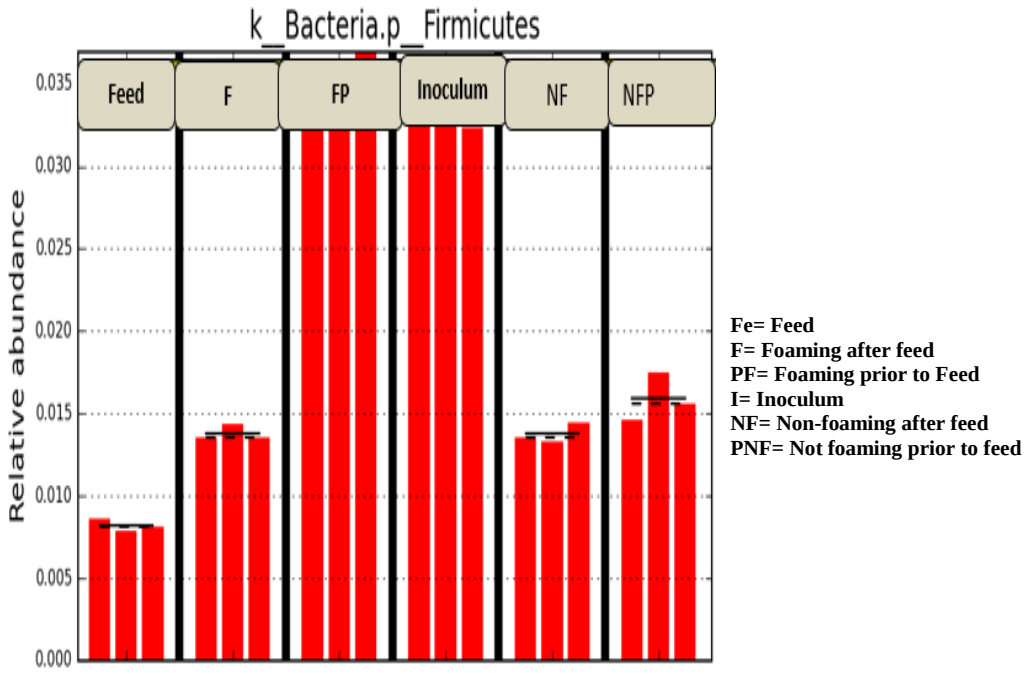


Figure 6-8: Relative abundance of firmicutes in the biological samples

As reported by R.M.W Ferguson et al. (2016), there were similarities in the bacterial community changes of two AD subjected to organic over load using Glycerol and FOG. Decreased in biogas methane content resulted to triple and double increase in Firmicutes for glycerol and FOG respectively. The same trend was observed in this experiment as there was a decrease in the relative abundance of the foaming digester when it was fed during the experiment. However, there was no significant change in the relative abundance of Firmicutes in the non-foaming digesters before and after the feed. Thus, increase in Firmicutes could be related to poor AD operation as they are mostly responsible for acidogenesis and acetogenesis, when they proliferate much more that the methanogens, it will lead to AD imbalance.

On a closer survey of the impact of the different microbial population there was a very unique display of relative abundance by the fibrobacter as shown in figure 6-9 which attracted further research. It could be observed that the Fibrobacter was found in abundance in the foaming and non-foaming samples after feed. This suggests that this type of bacteria has a very high growth potential as well as resilient. It highlights that its activities is not affected by system perturbation thus could be a very useful microorganism for most complex biotechnological process.

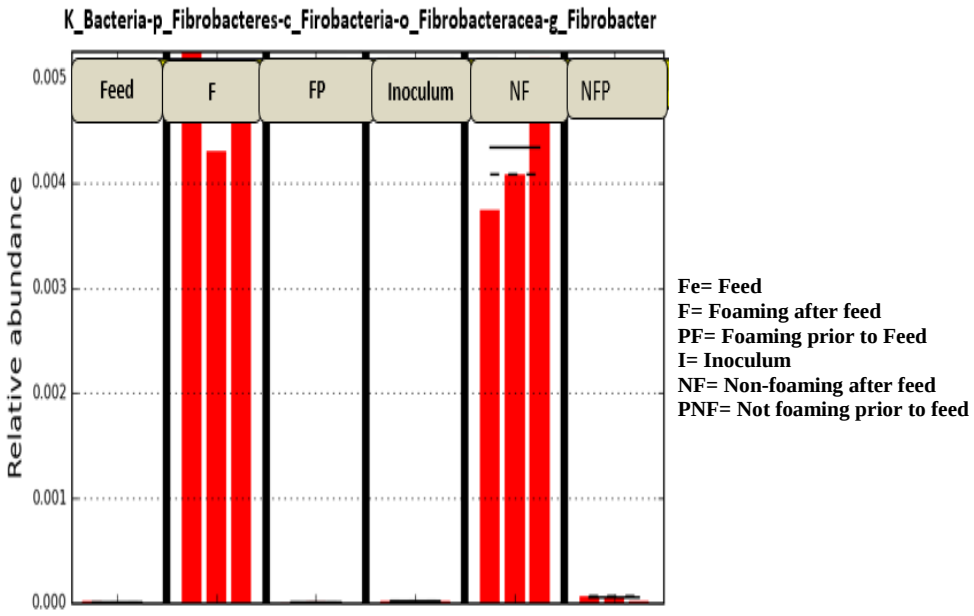


Figure 6-9: Relative abundance of Fibrobacter in the different biological samples

This is also supported by available literature on fibrobacter existence and impact in the hydrolysis of biomass (Ransom-Jones, et al., 2012; Kobayashi, et al., 2008; Ransom-Jones, et al., 2014). Fibrobacter was found to be typically adapted to the biodegradation of cellulosic plant. Ransom-Jones, et al. (2012) carried out 16S rRNA gene sequencing studies with the identification of Fibrobacteres in landfill sites, freshwater lakes and the termite hindgut attesting to a wider ecological range. It was concluded that since the Fibrobacters are constrained to effective cellulolytic degradation, they could be of considerable functional importance in the conversion of lignocellulosic biomass in the environment (Ransom-Jones, et al., 2012). However, the result of this analysis tends to support this proposal as well highlighting the fact that they may not actually be limited to cellulose degradation but could be very efficient for the hydrolysis of other form of biomass.

Furthermore, phylum actinobacteria representing the most prominent filamentous foam causing bacteria such as Norcadia amarae and Microthrix Parvicella had a very low relative abundance of 0.9% in the foaming and non-foaming AD indicating that the foaming occurrence in the AD studied was not triggered by the presence of filamentous bacteria.

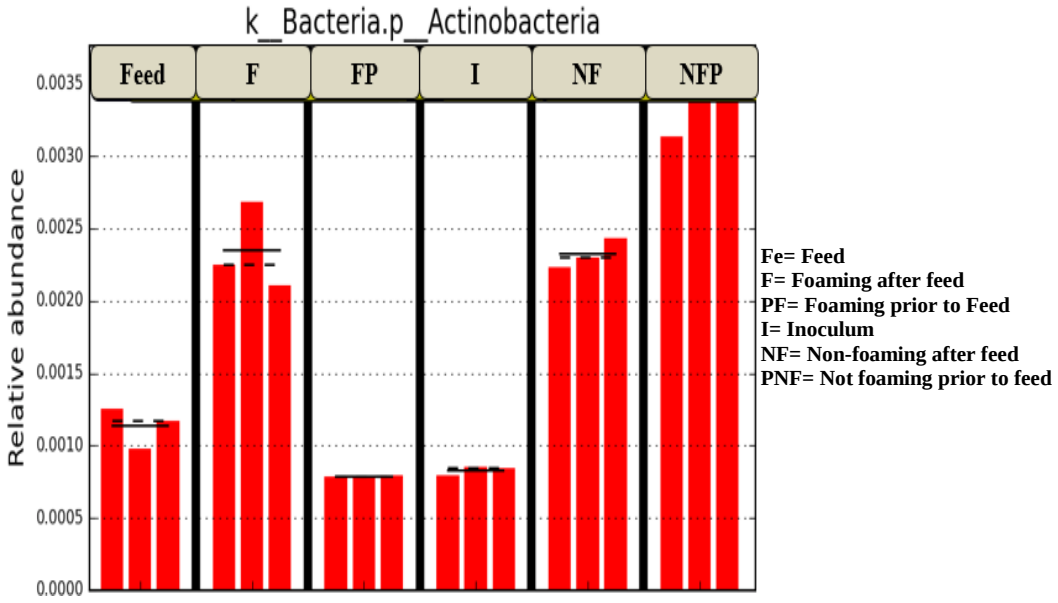


Figure 6-10: Relative abundance of Actinobateria in the different biological samples

The relative abundance of Actinobacteria for the samples were: 0.4% (Foaming before and after feed) 1.2% (Non-Foaming before and after feed), 0.6% (Feed) and 1.7% (Inoculum). It will be observed that the relative abundance of the Actinobacteria decreased during foaming occurrence thus signifying that foaming in this experiment is not due to filamentous microorganism.

Based on the preceding discussions and linking it to the key identified stages of anaerobic digestion process, key deduction could be reached as follows;

- Bacteria which are mainly Bacteriodietes were responsible for the hydrolysis of the complex substance
- The Firmicutes could be linked with the acidogenesis and acetogenesis phase of the anaerobic digestion process as could be judged from the greater abundance of firmicutes in the foaming digester while
- The Euryarcheota were responsible for the methanogenesis phase of the anaerobic digestion process.
- When comparing the foaming and non-foaming AD, the most abundant bacteria were not filamentous as there was no meaningful change in

the phylum actinobacteria (*Norcadia amarae* and *Microthrix Parvicella)* notable for foaming in AD.

In summary, the high accumulation of VFA in the foaming digester due to metabolomic analysis could be linked to the high level of hydrolysis, acidogenesis and acetogenesis that is taking place in the digester without a substantial number of methanogens to convert the product to methane. It could be observed that the abundance of bactriodetes in the foaming and non-foaming digesters were almost the same. On the other hand, the abundance of the firmicutes was decreased by half while the Euyarcheoata abundance doubled in the foaming sample compared to the non-foaming sample. This is essential to identifying the most significant stages in the anaerobic digestion process to effectively control. It is obvious that for the system to function properly there should be a balance between the abundance of microorganism responsible for acetogenic and methanogenic phase of the AD process.

6.3 Metabolomic analysis

To assess the metabolic variation and reproducibility of the samples, PCA was performed as shown in figures 6-10 for the set of experimental classes. The PCA of the metabolomics test was performed using 588 signals identified as likely metabolites attached as appendix 3. The PCA plot (figure 10) shows the unsupervised clustering of the samples. All experimental classes were clustered separately

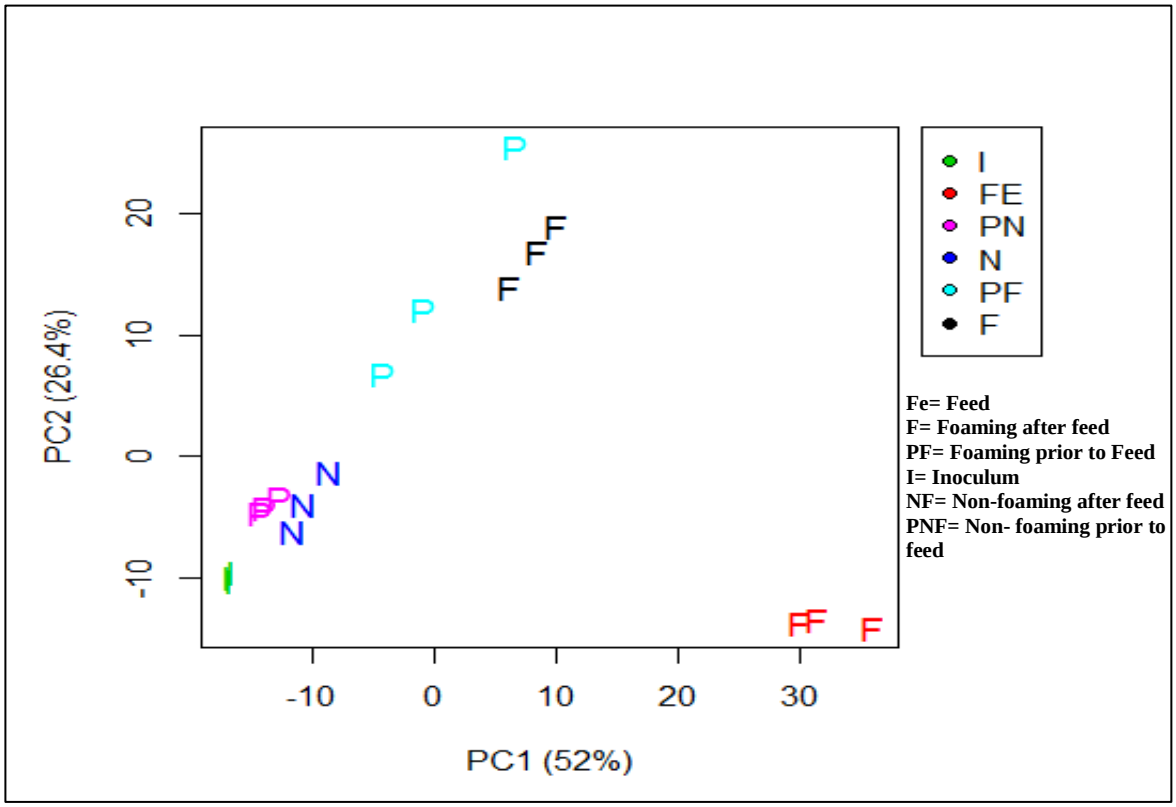


Figure 6-11: Plot of the first two principal components calculated for the experimental biological classes used in the metabolomics analysis:

Although a high variability of metabolite levels should have been expected in such complex samples, however, the results show that samples are highly reproducible. This highlights the stability of AD process when operated within the same conditions. The significance of this deduction informs the ability to adopt result from a specific AD experiment for other design and inference for ADs operating within the same condition, thereby supporting the development of predictive models for optimizing of AD process treating activated sludge using data driven techniques.

In addition, figure 6-10 depict that there is higher concentration of metabolites in the foaming samples before and after the feed. On average, the concentration of metabolite in the foaming sample was highest thus highlighting the fact that there has been accumulation of metabolites such as VFA in the digester. It could be observed that the feed sample had the lowest concentration of metabolites as the condition and microbial requirement to enable breakdown and generation of metabolites were not available.

As the inoculum was the same sample used to start up the experiment which have been in the storage for a long period of time at 4℃, thus, it could be observed that it seem much more like a technical replicate without much variations within the samples while exhibiting a lower concentration of metabolites compared to sludge sample from the active digesters.

In addition, profound differences was identified between foaming and non-foaming cultures, especially in terms of fatty acid and other lipid levels as shown in figure 6-11.

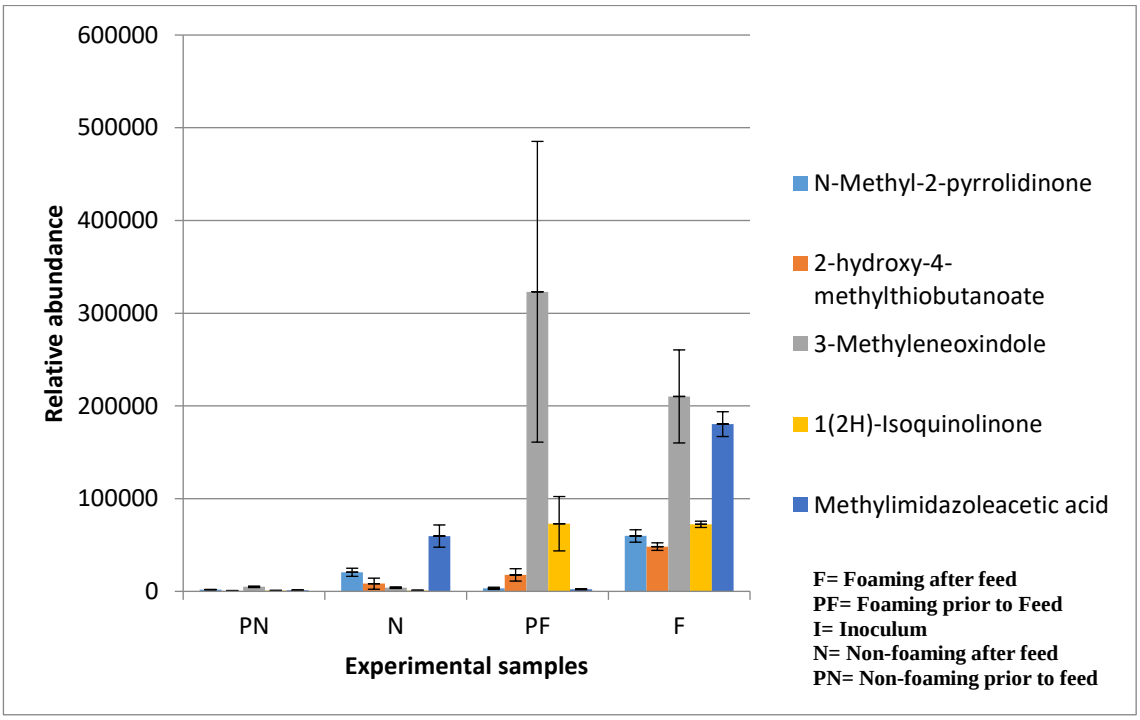


Figure 6-12: Typical comparison of different metabolites to estimate their presence in the different samples

The was potentially of great interest as the levels of many fatty acids were significantly and consistently higher in the foaming pre and post feed samples than in the non-foaming samples. This clearly offers new directions in analysis which require that the fatty acids content at the start of a fermentation should be ascertained to assist in efficient digester operation. The significant literature linking the proportions of volatile fatty acids to anaerobic digester quality (Aaring et al. 1995, Hori et al, 2006), and the literature implicating long chain fatty acids in digester failure (e.g. Angelidaki et al. 1992) and foaming (Gallart et al. 2002) are additional

evidence to explain why the foaming occurs in samples with higher levels of longer chain fatty acids.

Table 6-2: Metabolite in the post feed samples reflecting the high level of Lipid

Mass	RT	Formula	isomer	Putative metabolite	Confidenc	Map	Pathway	Maximum	PNF	NF	PF	F
Mass	RT	FORMULA	Iso	Putative metabolite	cor	Map	Pathway	intensity				
300.27	3.827	C18H36O3	27	[FA hydroxy(18:0)] 2S-hydroxy-octadecanoic	7	Lipids: Fatty Acyls	Fatty Acids and Conjugates	11725166	1.00	2.77	11.29	15.05
280.24	3.748	C18H32O2	84	Linoleate	5	Lipids: Fatty Acyls	Linoleic acid metabolism	11388443	1.00	1.80	3.27	3.98
100.05	13.6	C5H8O2	16	Tiglic acid	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	730798	1.00	20.39	2.07	64.77
282.26	3.726	C18H34O2	29	[FA (18:1)] 9Z-octadecenoic acid	7	Lipids: Fatty Acyls	Fatty acid biosynthesis	26867632	1.00	2.43	6.59	7.50
226.19	3.83	C14H26O2	12	(9Z)-Tetradecenoic acid	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	494027	1.00	2.73	5.42	8.68
254.22	3.774	C16H30O2	19	(9Z)-Hexadecenoic acid	6	Lipid Metabolism	Fatty acid biosynthesis	3340572	1.00	3.45	6.18	9.54
102.07	7.631	C5H10O2	5	Pentanoate	7	Lipids: Fatty Acyls	Fatty Acids and Conjugates	18508438	1.00	3.85	116.77	121.86
200.18	3.895	C12H24O2	10	Dodecanoic acid	8	Lipid Metabolism	Fatty acid biosynthesis	2203141	1.00	1.92	15.19	22.30
298.25	3.763	C18H34O3	46	2-Oxooctadecanoic acid	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	2152490	1.00	2.66	29.22	29.31
278.22	3.723	C18H30O2	38	[FA (18:3)] 9Z,12Z,15Z-octadecatrienoic acid	5	Lipids: Fatty Acyls	alpha-Linolenic acid	1101584	1.00	2.32	0.84	1.83
252.21	3.779	C16H28O2	16	[FA (16:2)] 9,12-hexadecadienoic acid	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	182687	1.00	4.65	2.25	6.44
268.24	3.775	C17H32O2	18	omega-Cyclohexylundecanoic	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	284969	1.00	2.53	6.70	8.66
228.21	3.813	C14H28O2	15	Tetradecanoic acid	8	Lipid Metabolism	Fatty acid biosynthesis	6397348	1.00	1.82	13.96	18.05
316.26	3.96	C18H36O4	19	[FA hydroxy(18:0)] 9,10-dihydroxy-octadecanoic	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	63142	1.00	3.28	11.79	17.80
310.29	3.69	C20H38O2	16	[FA (20:0)] 11Z-eicosenoic acid	6	Lipids: Fatty Acyls	Biosynthesis of unsaturated	394176	1.00	2.19	11.67	13.79
242.22	3.787	C15H30O2	9	[FA methyl(14:0)] 12-methyl-tetradecanoic	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	2186615	1.00	1.82	14.83	16.65
132.08	4.954	C6H12O3	14	[FA hydroxy(6:0)] 4-hydroxy-hexanoic acid	5	Lipids: Fatty Acyls	Fatty Acids and Conjugates	48080	1.00	8.60	3.75	39.13
304.24	3.789	C20H32O2	38	[FA (20:4)] 5Z,8Z,11Z,14Z-eicosatetraenoic acid	6	Lipids: Fatty Acyls	Fatty Acids and Conjugates	112735	1.00	5.81	1.82	13.43
172.15	4.039	C10H20O2	9	Decanoic acid	8	Lipid Metabolism	Fatty acid biosynthesis	448550	1.00	2.00	23.03	29.94

Table 6-2 present a non-exhaustive list of the first most changing putatively identified 19 fatty acids. It could be observed that there is a lot of Lipid and amino acid metabolism taking place in the digesters. This further support the fact that carbohydrate metabolism happens faster than protein and Lipid metabolism. This is essential to identifying the source of biodegradable material that should be effectively monitored to limit system perturbation especially if co-digestion of sewage sludge and other biodegradable waste stream is to be considered.

6.4 Biochemical Pathway

Every living organism undergo some form of metabolism that consists of many biochemical pathways through which substrate molecules are converted into specific products needed by the microorganism to thrive. The putative metabolites that were used in this metabolomics analysis are as shown in appendix 3. These putative metabolites make up the different product of each biochemical process. The biochemical process may be anabolic or catabolic. Anabolic pathways build complex products from less complex substrates while catabolic pathways breakdown complex products from less complex molecules to make energy available for biological processes. Thus, biochemical pathways essentially interact with one another so that cells have maximum flexibility in converting a wide array of substrates into specific molecules needed to thrive.

Thus a clear understanding of the different biochemical pathway going on in the AD and understanding the role the different microorganism play in achieving each step in the AD process could proffer new and more proficient ways of operating AD in the future.

6.5 Conclusion

The result of the metagenomics and metabolomics analysis enabled the identification of different microorganism and metabolite that were prevalent in the AD for the foaming and non-foaming samples. These results further support the outcome of the laboratory experiment on the impact of microbial population and metabolites on system perturbation. Although metagenomics analysis has been extensively applied in the past to analyse AD samples, this is the first-time metabolomics analysis is been carried out on AD sample thus

the result has widened the knowledge of AD process as well as confirming the factors responsible for AD perturbation of which a high concentration of VFA is a key factor.

Chapter 7 : Predictive modelling using artificial intelligence

In this chapter, the development, verification and validation of foaming preventive models were presented.

7.1 Variable selection for predictive modelling

A combination of the analysis from the laboratory experiment as detailed in chapters 4,5 and 6 formed the basis for selecting the following variables to be best suited for the predictive model;

1. To develop a more robust and versatile model, organic loading rate was chosen over volume of sample especially in scenarios where the volatile solid concentration of the feed sludge varies. However, if the volatile solid concentration of the feed does not vary, then the volume of sample will be a more easily determined and useful variable for the predictive model.

2. As discussed in section 5.4, VFA in the feed was selected though it does not exhibit a high or significant correlation with any of the variables except FI index in the feed. Thus, the decision to still include it as an input variable was informed by the synchronised occurrence of the peak VFA concentration and the onset of foaming. As noted previously in section 5.1.3, no other variable exhibited this pattern of behaviour with foaming.

3. Maximum volume of foam as a percentage of the digester size as output variable. In choosing the best output for the predictive model, maximum volume of foam recorded in the digester represented as a percentage of digester volume was selected as against the additional digester volume occupied by foam as it is deemed that the result will be more useful to the operators of anaerobic digester. This is because the operators will find more value in using a facility that will enable them to avert the commencement of foaming rather than preparing measures to overcome the effect of nuisance foaming.

7.2 Stochastic data generation

Synthetic data generation for the purpose of extending the data records was limited to the maximum volume of foam (MVF) as output and the two input variables that significantly influence the foaming as established in section 7.1, namely, organic loading rate (OLR) and Volatile Fatty Acid (VFA).

Additionally, as shown in section 5.3, the data collected during the experiment for the selected variables were normally distributed; hence the adoption of a standardised normal variate with a mean of zero and variance of 1, during synthetic data generation using Markov lag-one auto regressive AR (1) as discussed in section 2.8.2.3 and 3.2.7 using equations 2-7a, 2-7b and 2-7c.

The Markov lag-one AR (1) model (equation 2-6) with parameters mean, standard deviation and auto correlation coefficient estimated from the observed data (see table 7-1) was then used to generate 3100 data points for each of the three variables for subsequent analyses. In running the data generation model, the initial value was assumed to be equal to the observed mean value. Hence, only the latter 3000 data points were used; the first 100 data points were discarded because the data have been influenced by the assumption regarding the model initialisation

The estimated parameters of the generated data are also presented in table 7-1 together with the respective errors relative to the observed parameters.

Studying the output of the stochastic data generation as shown in table 7-1, it could be observed that the parameters mean, standard deviation and lag one auto regressive coefficient were effectively preserved in the synthetic data especially for VFA and OLR.

The large error in the mean for MVF is accounted for by the need to expose the model to various range of foaming following the introduction of values lower than 36% and higher than 80% which were not within the data collected from the experiment.

Table 7-1: Percentage variation in mean for observed and synthetic data generated using Markov model and adjusted Markov model.

Parameters	OLR			VFA			MVF		
	Measured Kg/m^3d	Synthetic Kg/m^3d	Error %	Measured mg/l	Synthetic mg/l	Error %	Measured %	Synthetic %	Error %
Mean	3.02	3.15	-4.60	145.79	144.38	0.96	24.58	31.22	27.05
STDev	2.23	1.98	11.48	48.88	49.21	-0.68	29.40	22.49	23.51
Lag one coefficient	0.92	0.90	1.82	0.43	0.44	-1.98	0.96	0.93	2.49

However, as the lag one auto regressive coefficient is preserved this support the fact that the figures are relevant and are essential to making the model robust and more efficient.

7.3 Modelling in MATLAB

The synthetic data generated was divided into two; 2500 data points were set aside for modelling, testing and checking while 500 data points were used to validate the model as discussed in the different section for modelling the data. This is because, the randomness introduced in generating the synthetic data allowed for a more strengthened data possessing a comprehensive overview of the anaerobic digestion process and foaming occurrence.

7.3.1 Fuzzy logic modelling

Fuzzy logic permits the application of human knowledge of the system to be modelled in order to avert the constraint of depending solely on the analysis of numbers. Thus, applying the basic concept underlying fuzzy logic which incorporates linguistic variable allows the model to behave closely to human intuition permit the exploitation of tolerance for imprecision thereby lowering the cost of solution.

7.3.1.1 Mamdani Fuzzy Logic Modelling

As discussed in section 2.8.2.4.2.1; using the Mamdani fuzzy logic requires that the inputs and outputs are partitioned into fuzzy regions. The complexity encountered in the application of the Mamdani model stems from the fact that the number and the shape of the membership functions are difficult to determine. In addition, the number of fuzzy rules increases dramatically as the number of input variable increase.

Thus, a priori knowledge about the system becomes very useful in selecting the structure of the model. Consequent to extensive knowledge acquired about the system to be modelled through laboratory experiment and participating in the full-scale operation of anaerobic digesters, expert knowledge was applied in developing the Mamdani model.

7.3.1.1.1 Selecting the membership function.

Using MATLAB Fuzzy Logic tool box in this study, expert knowledge was backed up with series of trial and error to determine the membership function that result to least error in predicting the initial onset of foaming. Such model ensures a high correlation between the actual and predicted value. Another important factor that requires great consideration in developing Fuzzy logic model is to ensure that the rules are parallel, as this will allow the smooth flow of logic from points of where the characteristics of the system modelled is dominated by either one rule or another.

To ensure that the selection was fair, the same number of input (2), input membership function (3), output (1), output membership function (3) were applied to the different form of membership function selected for this trial as shown in table 7.2. The input membership functions were organic loading rate (OLR) and volatile fatty acid concentration in the feed (VFA) while the output was maximum volume of foam (MVF) observed in the digester as a percentage of digester volume. The membership functions used were triangular (Tr), trapeizoidal (Tr), GBell (GB) and Gaussian (G). The same form of membership function was used for the input and output variables while the

shapes were varied over ten times before selection of the model presented as the best representative of the sect. as shown in table 7-2.

The ability of MFL developed to accurately predict foaming occurrence were evaluated using the 500 data points set out for model validation and the result presented in table 7-2. This was essential to enable an easy comparison of result with other modelling techniques.

Table 7-2: Structure and performance of Mamdani Fuzzy logic in predicting maximum volume of foam

Model	Correlation (%)	MAD	MSE	RMSE	Max	Min	Mean
Measured data	NA	NA	NA	NA	80	0	24.58
Triangular	61.05	11.29	200.63	14.14	85.12	12.36	36.84
Trapezoidal	42.17	13.42	283.26	16.83	86.32	11.23	38.31
GBell	60.28	11.41	200.37	14.16	87.31	13.46	40.73
Gaussian	57.38	12.28	223.94	14.97	89.69	11.57	37.22

It could be observed from table 7-2 that although a lot of effort was put in tuning the Mamdani fuzzy logic models (MFL) based on good understanding of the collected data and the anaerobic digestion process, MFL could not effectively predict foam formation in anaerobic digestion process. This is mainly due to the complexity of the data and the operating condition.

MFL model using triangular membership function was the most representative of the foaming occurrence with a correlation of 61.05% and RMSE of 14.14. It is worth mention that the ability of MFL model using triangular membership function to give a better prediction for the data collected could be related to the simplicity in determining the parameters and the ability to easily manipulate it to achieve a better output. The MFL model with GBell membership function also performed close to the model developed with triangular membership function with a correlation coefficient of 60.28% and RMSE of 14.16.

Plotting a trend of the MFL models as shown in figures 7-1 to 7-4; it could be observed that although the models were not very efficient at accurately predicting foaming occurrence, they were able to predict the onset of foaming as relate to the result generated from the experiment.

During the experiment, initial foaming occurrence was 36% of digester volume, thus, it could be viewed from figure 7-1 that all foaming occurring within the 36% mark was effectively modelled such that most of the variation of the MFL model was for the higher and lower values introduced by the synthetic model which may not have been sufficiently captured in developing the model.

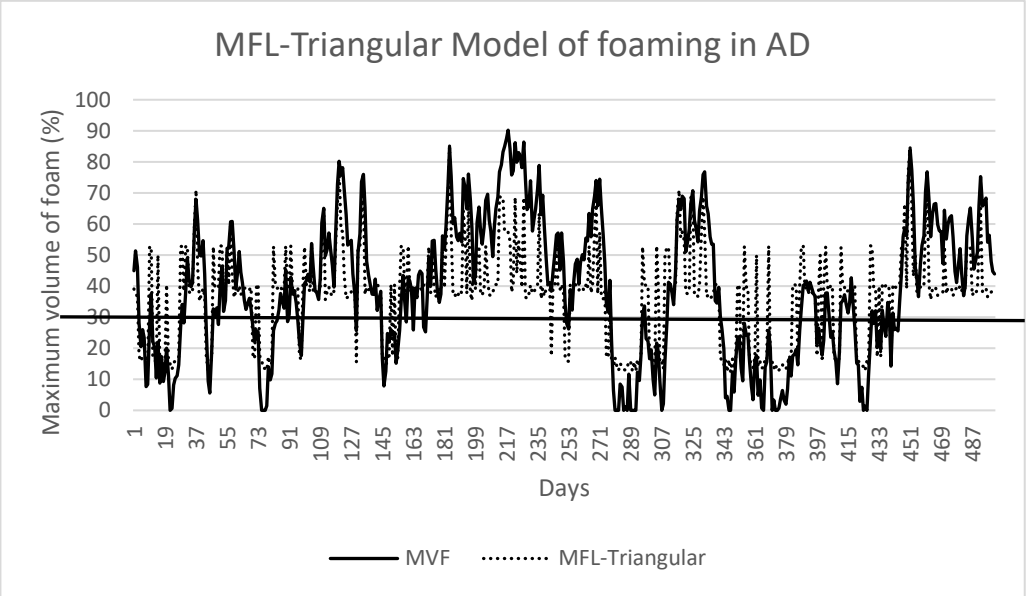


Figure 7-1: MFL-Triangular Model of foaming in AD

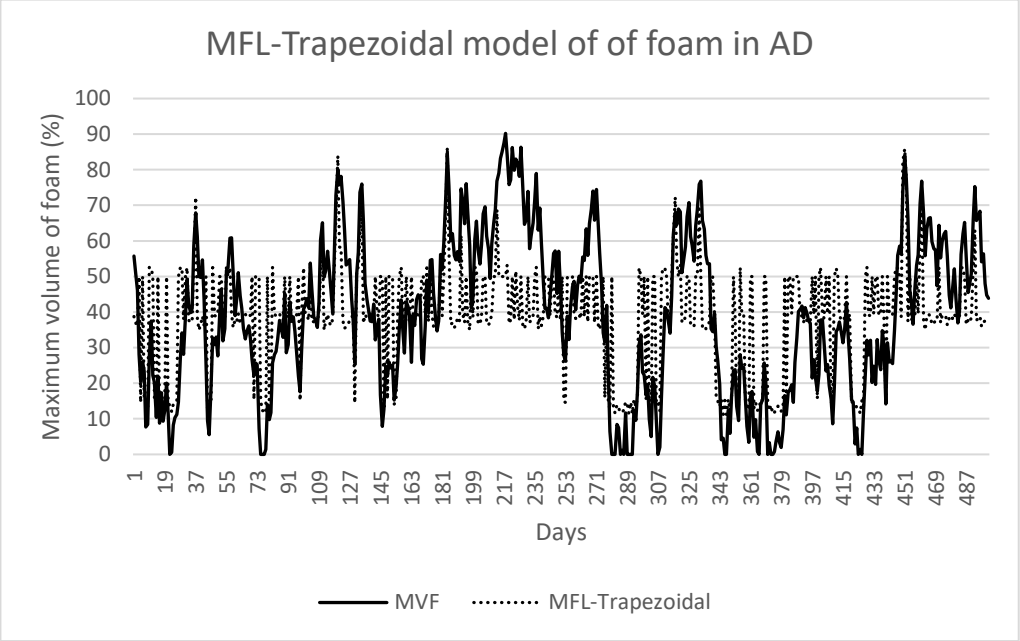


Figure 7-2: MFL-Trapezoidal Model of foaming in AD

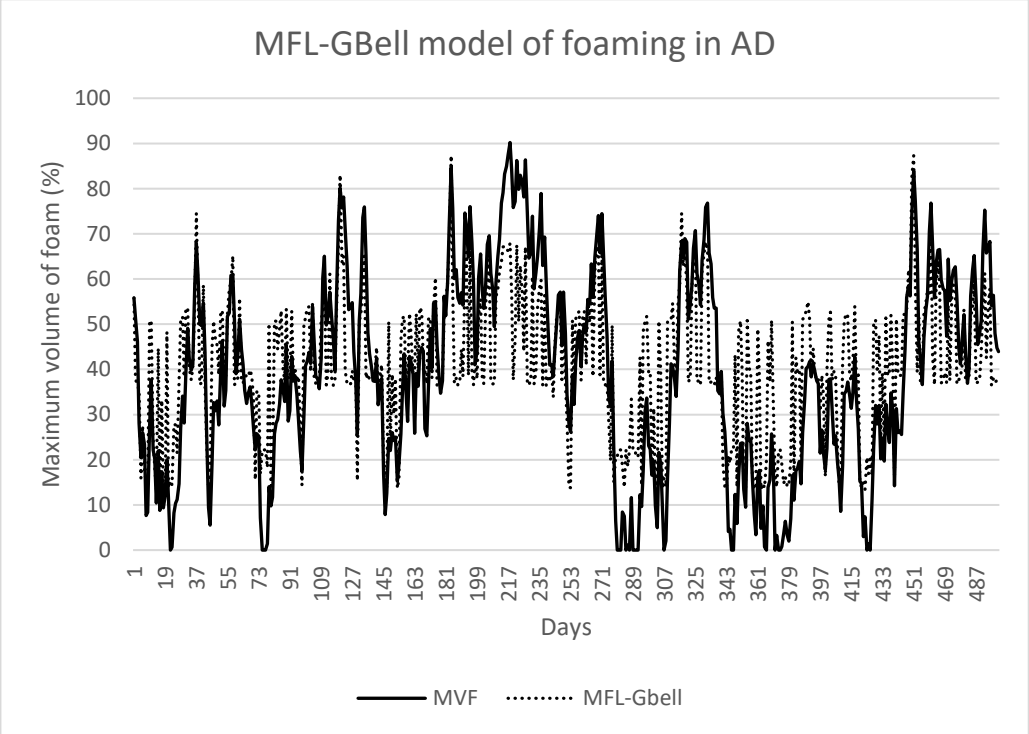


Figure 7-3: MFL-GBell Model of foaming in AD

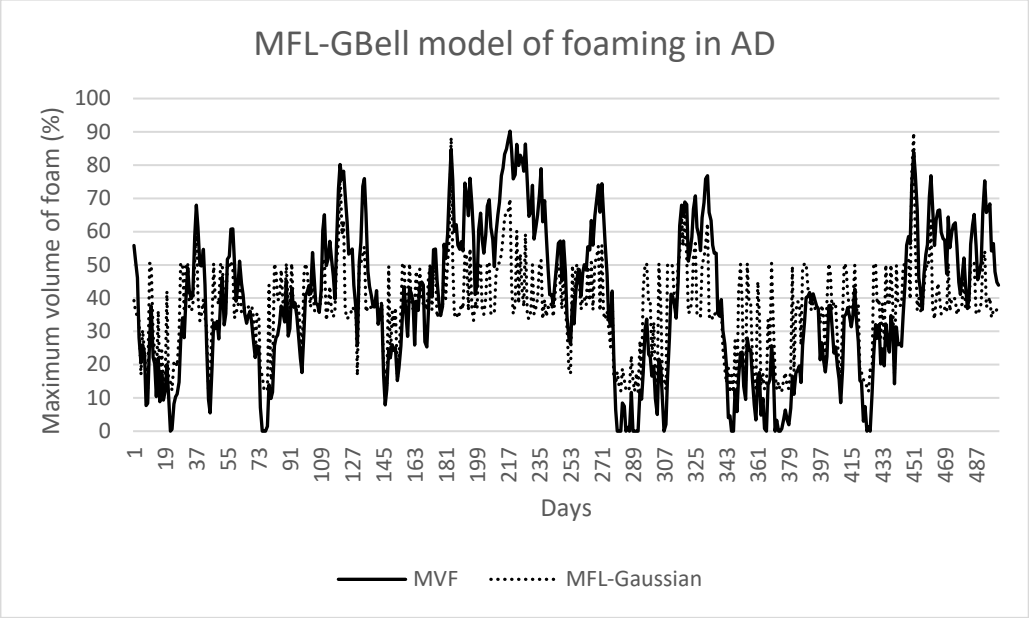


Figure 7-4: MFL-Gaussian Model of foaming in AD

7.3.1.1.1 Discussion on the fuzzy logic model

Since MFL-triangular model yielded the best result, the structure of that model will be discussed further. Firstly, the extents to which the inputs belong to each of the appropriate fuzzy sets through the membership functions were determined and this is known as fuzzification of input. For the input variables, the fuzzy sets for OLR were low, medium and maximum while for VFA, they were Low, average and high as shown in figure 7-5. The output fuzzy sets applied in the implcation were no foaming,foaming and nuisance.

Based on these fuzzy sets, fuzzy logic If-Then rules governing the model were developed leveraging on extensive knowledge of the system to form the following linguistic sets;

1. If OLR is minimum (antecedent) and VFA is low (antecedent), then, MVF is nil (consequent) with a weight of 1.
2. If OLR is minimum and VFA is average, then, MVF is ni with a weight of 0.8.
3. If OLR is medium and VFA is average, then, MVF is nil with a weight of 0.5

4. If OLR is medium and VFA is average, then MVF is foaming with a weight of 1.

5. If OLR is maximum and VFA is high, then MVF is foaming with a weight of 1.

6. If OLR is maximum and VFA is high, then MVF is nuisance with a weight of 1.

7. If OLR is maximum and VFA is average, then MVF is nuisance with a weight of 1.

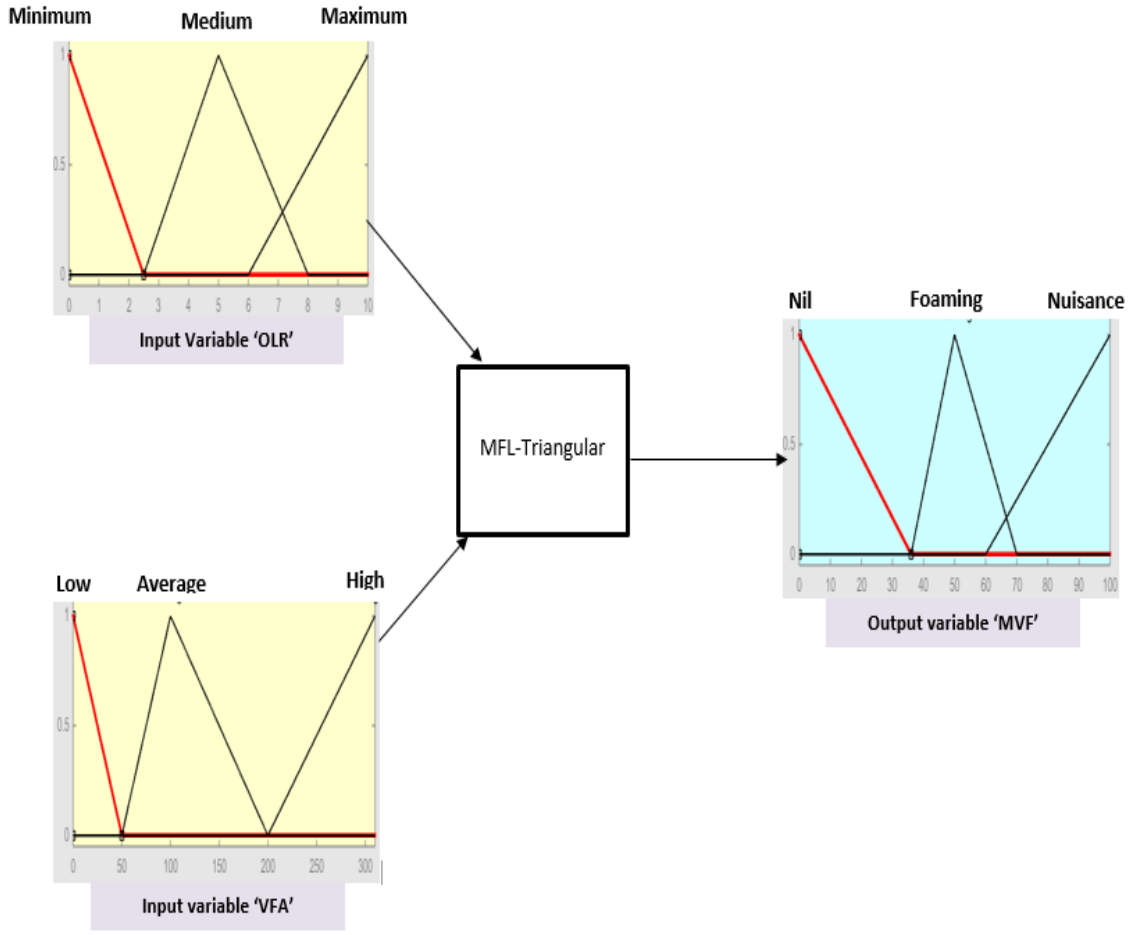


Figure 7-5: MFL-Triangular membership functions

With the input fuzzified, the degree to which each part of the antecedent is satisfied for each rule could be established and applied to the output function. However, as the rules were generating more than one membership value from fuzzified input variable, the fuzzy logical operator was applied to obtain one number that represent the result of the antecedent for that rule. The logical

operators applied to all the model developed in this study for 'AND' method is minimum while for 'OR' method is maximum.

The weight for the rules were varied prior to effecting the 'AND' implication method (minimum) on the value obtained from the antecedent, thereby truncating the output fuzzy set to generate the consequent. Since implication is carried out for each rule, the resultant consequents were combined into a single fuzzy set for the output variable using the maximum aggregation method. The aggregated output fuzzy set is defuzified to a single number by applying the centroid deffuzification method. Figure 7-7 is an illustration of the fuzzy inference process for MFL-Triangular.

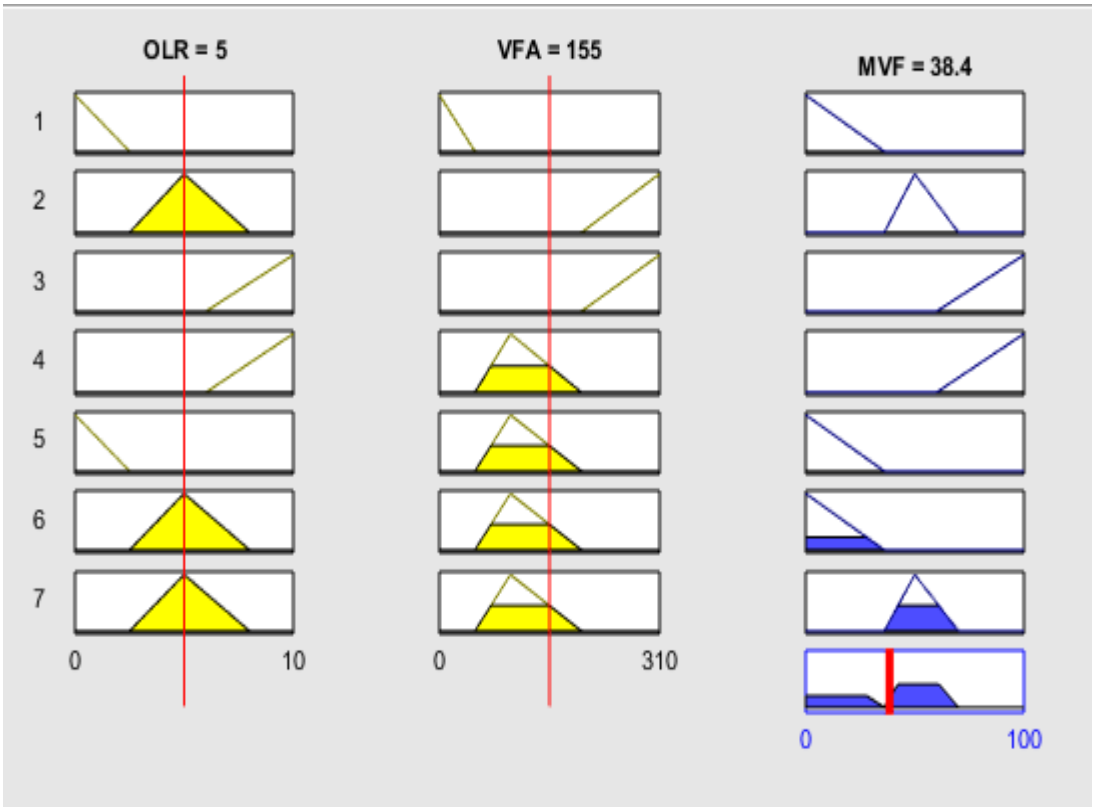


Figure 7-6: Fuzzy inference model for TpTrTr

Figure 7-4 is a graphical representation of the Mamdani TpTrTr and could be a very useful tool in understanding a lot about the model at a glance, thus could be very useful

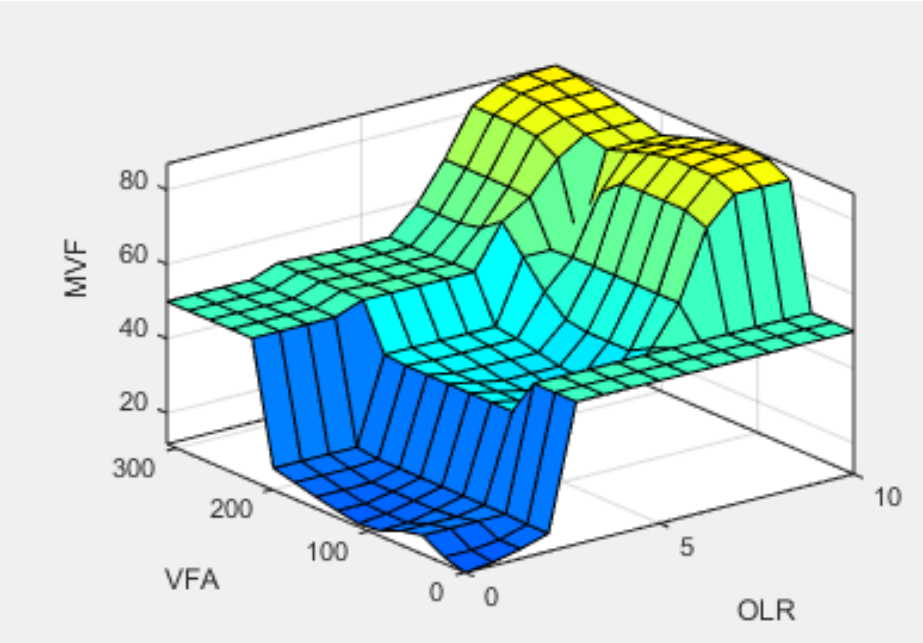


Figure 7-7: Graphical representation of MFL-Triangular model

In conclusion, the best four Mamdani models with a higher predictive performance were selected and presented in this report from various model developed for predicting foaming in anaerobic digester. The four models selected were able to predict foaming within the range of correlation from 61.05% to 57.38%.

A line plot of the measured and predicted foaming values as shown in figures 7-1 to 7-4 were very efficient at predicting the onset of foaming. However, it was difficult to precisely predict the continuous trend of foaming in the reactors. This is as a result of the erratic conditions that develops in the reactor once foaming commence of which the synthetic data was able to efficiently capture.

On the other hand, foaming occurrence constitute nuisance to the anaerobic digester operators thus it is not an issue of keen interest to be monitored rather should be mitigated. With due consideration to the fact that anaerobic digester operator's primary challenge is to predict the onset of foaming and put every measure in place to stop it from developing further, MamdaniTraining model has proved itself capable of doing that.

7.3.1.2 Neural Network Fuzzy Inference System (NNFIS)

NNFIS model uses a Sugeno method of fuzzy inference. It is similar to the Mamdani method as the first two parts of the fuzzy inference system (FIS), fuzzifying the inputs and applying the fuzzy operator, are precisely alike. The major difference between Mamdani and Sugeno is that the Sugeno output membership functions are either linear or constant.

In this study, grid partition was choosen over subtractive clustering for generating a single-output sugeno type FIS. The optimisation method used was a hybrid method. This involves applying backpropagation for the parameters associated with the input membership functions, and least squares estimation for the parameters associated with the output membership functions (Beale, et al., 2015). To access the model performance, the training error was determined by calculating the difference between the training data output value, and the output of the fuzzy inference system corresponding to the same training data input value recorded as the root mean squared error (RMSE) of the training data set at each epoch.

However, one of the greatest challenges in using NNFIS just like in any other fuzzy logic system is the ability to choose the most suitable membership function that will produce the best prediction. To ensure that this is carried out effectively and achieve a fair comparison with MFL, the four types of membership function (Triangular, Trapezoidal, GBell and Gaussian) used in MFL modelling were adopted. Out of the 2500 data points set aside for model training, 1000, 500 and 500 data points were uploaded for training, testing and checking the model respectively.

The same number of input membership function (3) was applied during the modelling. The epoch and output function (linear or constant) were varied over several model training times until the best model were identified and presented in table 7-3. The selection was based on monitoring the testing and checking error such that the training is ceased once the checking error start to increase which normally happen in most cases when the training error remains the same even with further increment in the number of Epochs. As shown in table 7-3, the Gaussian membership function had the least error.

Epochs	Training Error-RMSE									
	Training Error-RMSE									R^2
	3	10	20	30	40	50	70	80	90	
NNFIS-Triangular										
Constant	4.54	4.53	4.49	4.45	4.45	4.45				**94.26**
Linear	4.44	4.43	4.43	4.43						**94.30**
NNFIS-Trapezoidal										
Constant	8.98	8.67	8.09	7.24	6.22	5.27				**91.60**
Linear	4.44	4.44	4.36	4.43	4.43					**94.09**
NNFIS GBell										
Constant	5.92	5.68	5.26	4.84	4.57	4.51	4.49	4.46	4.4	**94.26**
Linear	4,41	4.41	4.41	4.40	4.38	4.39	4.39	4.39		**93.97**
NNFIS Gaussian										
Constant	4.79	4.68	4.55	4.48	4.47	4.47	4.47			**94.20**
Linear	4.42	4.41	4.41	4.41	4.41	4.41	4.41	4.4		**94.04**

It could be observed that the NNFIS model with the highest correlation was the NNFIS-Triangular-L while the NNFIS-Trapezoidal-C had the least performance. This follows with the same trend as was observed for the MFL model. Hence, these results support the fact that adequate expert knowledge was applied in developing the MFL model.

The training error was the same for all the constant output membership function at a figure of 4.43 except for the NNFIS-GBell model that ended with RMSE of 4.39, although with a lower correlation of 93.97. Based on the result as shown in table 7-3, the NNFIS models with the higher correlation were used for model validation and the results shown in tale 7-4 and figures 7-9 to 7-13.

Table 7-3: Validation of NNFIS model

NNFIS Model	Correlation (%)	MAD	MSE	RMSE	Max	Min	Mean
Measured data	NA	NA	NA	NA	90.21	0	39.18
Triangular-L	94.30	4.34	33.13	5.76	88.13	-1.08	37.06
Trapezoidal-L	94.09	4.40	33.80	5.81	87.83	-2.13	37.1
GBell-C	94.26	4.35	33.15	5.76	87.67	-0.52	37.08
Gaussian-C	94.20	4.37	33.50	5.79	87.73	-0.72	37.07

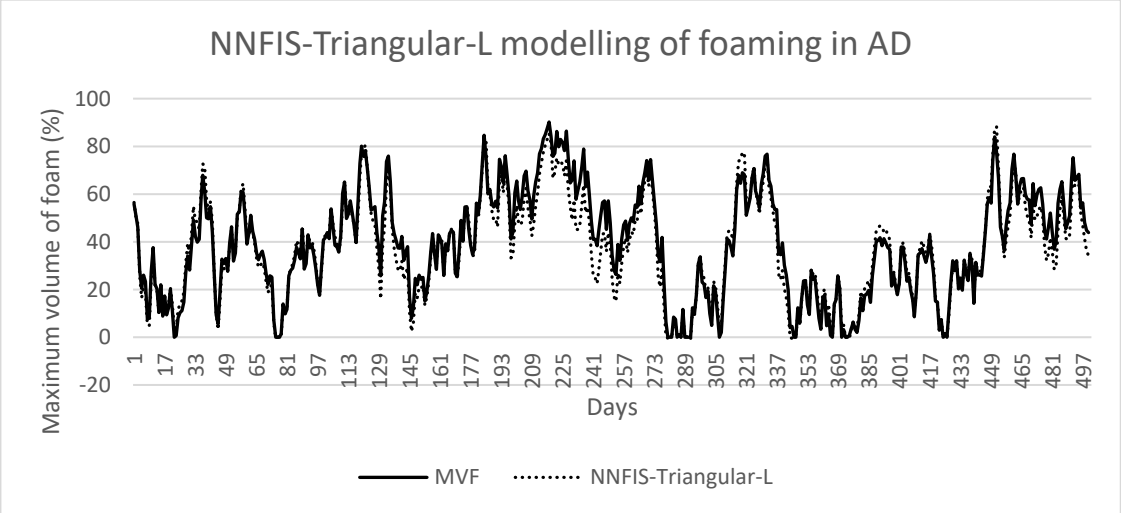


Figure 7-8: NNFIS-Triangular-L modelling of foaming in AD

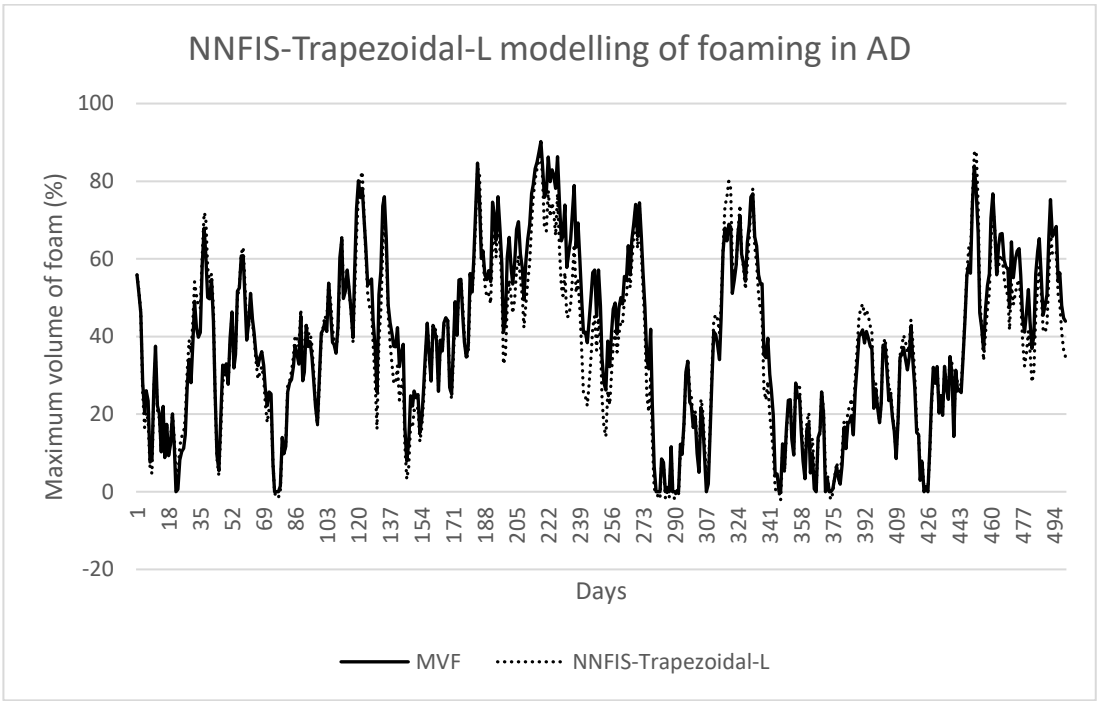


Figure 7-9: NNFIS-Trapezoidal-L modelling of foaming in AD

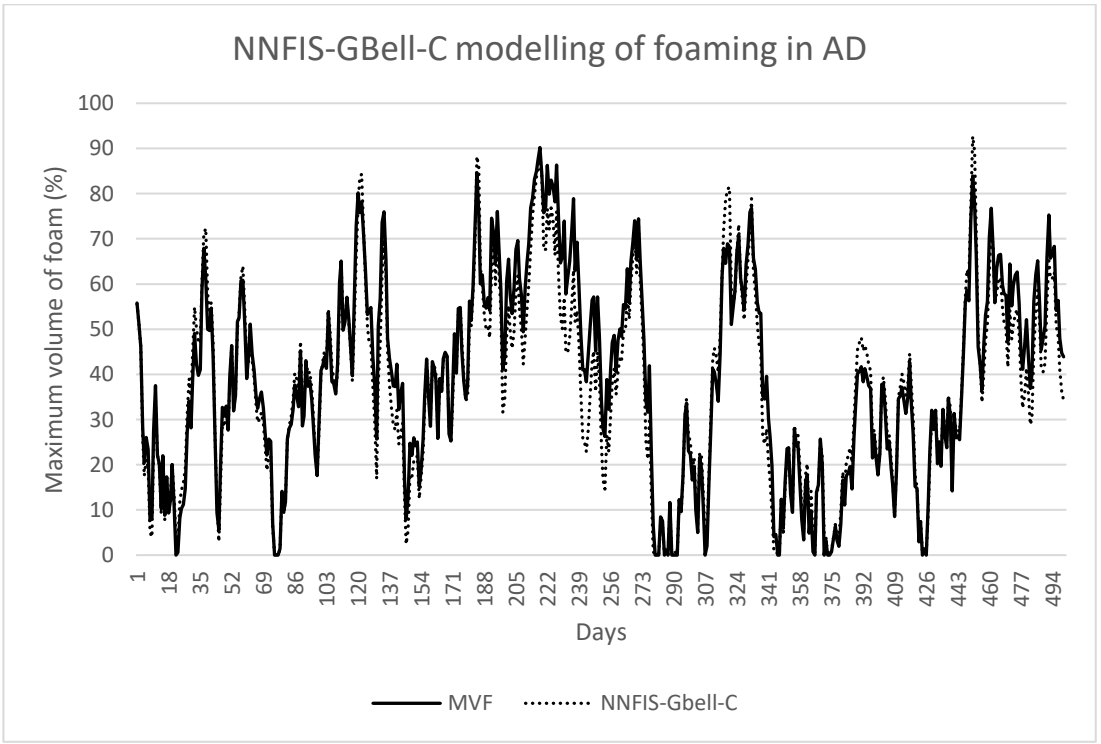


Figure 7-10: NNFIS -GBell-C modelling of foaming in AD

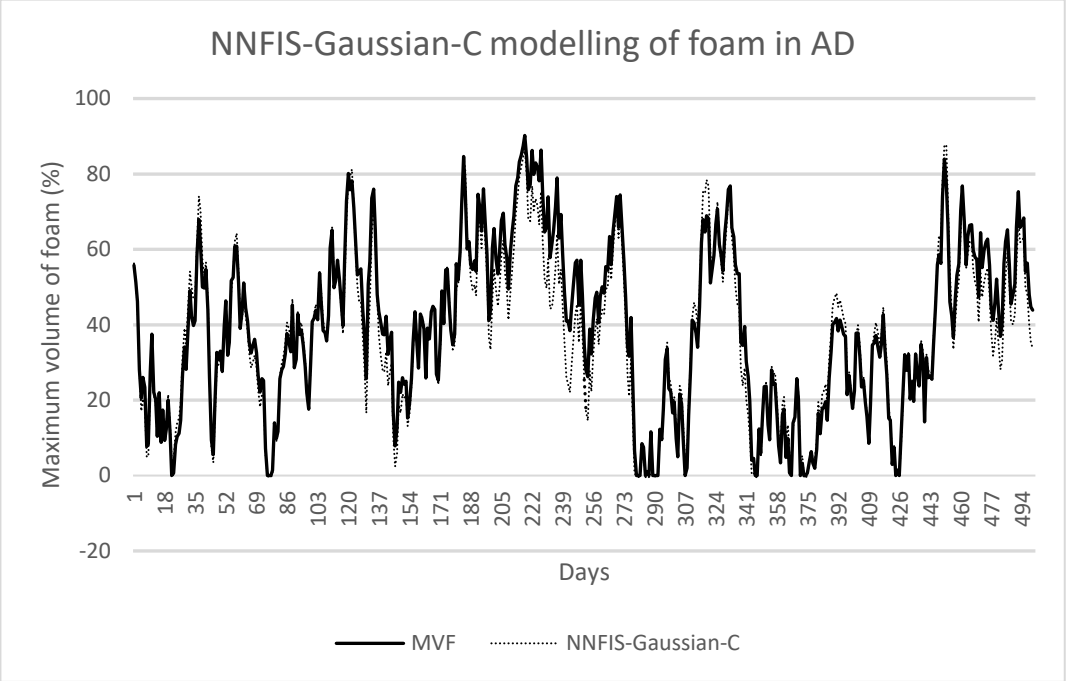


Figure 7-11:NNFIS-Gaussian-C modelling of foam in AAD

7.3.1.2.1 Discussion of NNFIS models

Further discussion on the NNFIS model will be done using the NNFIS-Triangular-L as it had the lowest RMSE and correlation with the measured data (Synthetic data) during model training. It could be observed from table 7-5 that the total number of parameters for each model was far less than the data applied in the model, meaning that this essential requirement for fuzzy logic modelling was met.

Table 7-4: Various Gaussian model developed changing the number of membership function.

Model	Number of membership functions in each input (N_{mf})	Number of linear Parameters P_1	Number of non-linear parameters P_2	Total number of parameters $P=P_1+P_2$	Number of Fuzzy Rules $N_{fr}=(N_{mf})^{Nin\ put}$
NNFIS-Triangular-Linear	3	27	12	39	9

Where: $P_1=N_{fr}*(N_{input}+1)$; $P_2=N_{input}*N_{mf}*2$; $N_{fr}=(N_{mf})^{Ninput}$

As shown in figure 7-12, the logical operators available for use in NNFIS were ‘and’, ‘or’ and ‘not’.

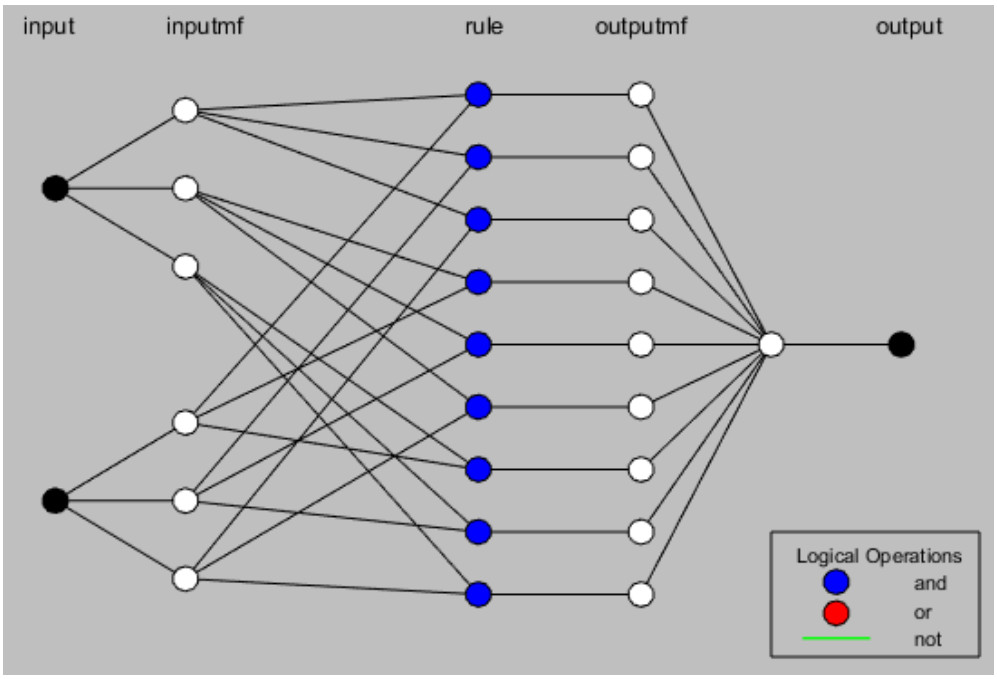


Figure 7-12: Structure of 2 inputs(3 membership functions); 1 output (Constant/Linear), NNFIS model

The ‘and’ (product) method was used for the nine rules developed for the NNFIS as shown in figure 7-12. With the input fuzzified by applying the rules, the degree to which each part of the antecedent is satisfied for each rule could be established and applied to the output function.

However, as the rules were generating more than one membership value from fuzzified input variable, the fuzzy logical operator was applied to obtain one number that represent the result of the antecedent for that rule. The logical operators applied to the entire model developed in this study for ‘AND’ method is product while for ‘OR’ method is probability.

The weight for the rules were varied prior to effecting the ‘AND’ implication method (minimum) on the value obtained from the antecedent, thereby truncating the output fuzzy set to generate the consequent.

Using the implication method (Product), the implication is carried out for each rule with the resultant consequents combined into a single fuzzy set for

the output variable using the maximum aggregation method. The aggregated output fuzzy set is defuzified to a single number by applying the 'wtaver' deffuzification method.

NNFIS Model	Correlation (%)	MAD	MSE	RMSE	Max	Min	Mean
Measured data used for validation (500)	NA	NA	NA	NA	90.21	0	39.18
Triangular-L	94.30	4.34	33.13	5.76	88.13	-1.08	37.06
Trapezoidal-L	94.09	4.40	33.80	5.81	87.83	-2.13	37.1
GBell-C	94.26	4.35	33.15	5.76	87.67	-0.52	37.08
Gaussian-C	94.20	4.37	33.50	5.79	87.73	-0.72	37.07

In conclusion, the NNFIS performance in predicting foaming in AD was better than the MFL models. With up to 94% correlation achieved by the different models of NNFIS in this study, it is evident that NNFIS is achieving a better result by combining neural network computing ability with fuzzy logic

7.3.2 Neural Network Model

Neural networks models have played significant role in the development of models for complex environmental systems (Cinar, 2005). As a form of computer modelling inspired by how the biological nervous system function, neural networks are non-linear statistical data modelling tools that models complex relationships between inputs and outputs or identify patterns in data (Rabee & Adebayo, 2009). Thus, the network acquires knowledge about the data through a learning process and store the knowledge by means of interconnections between elements of the network (Arbib, 2003).

One of the major constraints in the use of neural networks is the ability to specify the number of hidden neurones that forms part of the network architecture. Hence, when the neurons are so few, it can lead to under fitting while too many neurons can contribute to overfitting, in which all training

points are well fitted, but the fitting curve oscillates wildly between these points.

Thus in this study, training were carried out ten times before accepting any neural network for further data processing. To ensure that the training method selected suffice for the predictive model using the available data, Levernberg, Bayesian regularisation and Scaled conjugate were developed using three different hidden neurones of 3, 10 and 15.

The idea to limit the model development to these set of hidden neurones stemmed from the fact that for each of the training methods there was not a significant variation in the model error between the hidden neurones used. Thus, for the data under consideration, an increment in the number of hidden neuron does not result in much variation in the model efficacy. The data used for the model was the same as that used for the NNFIS model which is a total of 2500 data points divided for training (2000), validation (250) and testing (250). It is worth mentioning that this validation was only used for the training process as the 500 data points set aside for validation was eventually used to test each of the models This was essential to effectively comparing the result of the NN, NNFIS and MFL.

The result of the NN modelling for the different algorithm and variation in the number of neurone is presented in table 7-5.

Table 7-5: Result of the NN modelling using different algorithm and neurones

NNFIS Model	**MSE**			**Correlation (%)**		
No of Neurones	5	10	15	5	10	15
Levenber Marqudart						
Training	21.511	21.05	21.587	97.8	97.8	97.8
Test	21.36	22.80	22.655	97.4	97.50	97.7
Validation	25.126	23.29	19.79	97.2	98.07	97.7

Bayesian Regularisation						
Training	21.910	21.271	21.686	97.8	97.8	97.8
Testing	22.592	23.890	24.561	97.4	97.6	97.3
Validation	-	-	-	-	-	-
Scaled Conjugate						
Training	24.291	23.920	34.223	97.5	97.4	96.5
Testing	21.326	18.313	37.68	97.7	98.2	95.7
Validating	25.937	20.297	36.855	97.5	98.1	97.0

It could be seen from table 7-6 that the correlation value during the training for Levenberg Marqudart (NNLM) and Bayessian regularisation (NNBR) models were 97.8% which reduced for the testing and validation. On the other hand, the correlation for the Scaled conjugate model (NNSC) was lower for the training but increased during the validation. This further reflect that the Scaled Conjugate model tend to perform better for training large networks as discussed in section 2.8.2.4.1.1 while Bayesian regularization training produce better generalization.

The best performing model for the different algorithm as shown in table 7-5 were the 10 hidden neurones.The data summarised in table 7-6 show that the neural network models behaved same when presented with the same validation data (500) with a correlation value of 94.35% which is slightly higher than the 94% correlation achieved by NNFIs-Triangular-L model. Hence the figures 7-13 to 7-15 were the same for the different algorithms.

Table 7-6: NN model validation result

NNFIS Model	Correlation (%)	MAD	MSE	RMSE	Max	Min	Mean
Measured data	NA	NA	NA	NA	90.21	0	39.18
Levenberg Marqudart	94.35	4.29	30.19	5.49	91.15	-1.29	37.79
Bayesian Regularisation	94.35	4.29	30.19	5.49	91.15	-1.29	37.79
Scaled Conjugate	94.35	4.29	30.19	5.49	91.15	-1.29	37.79

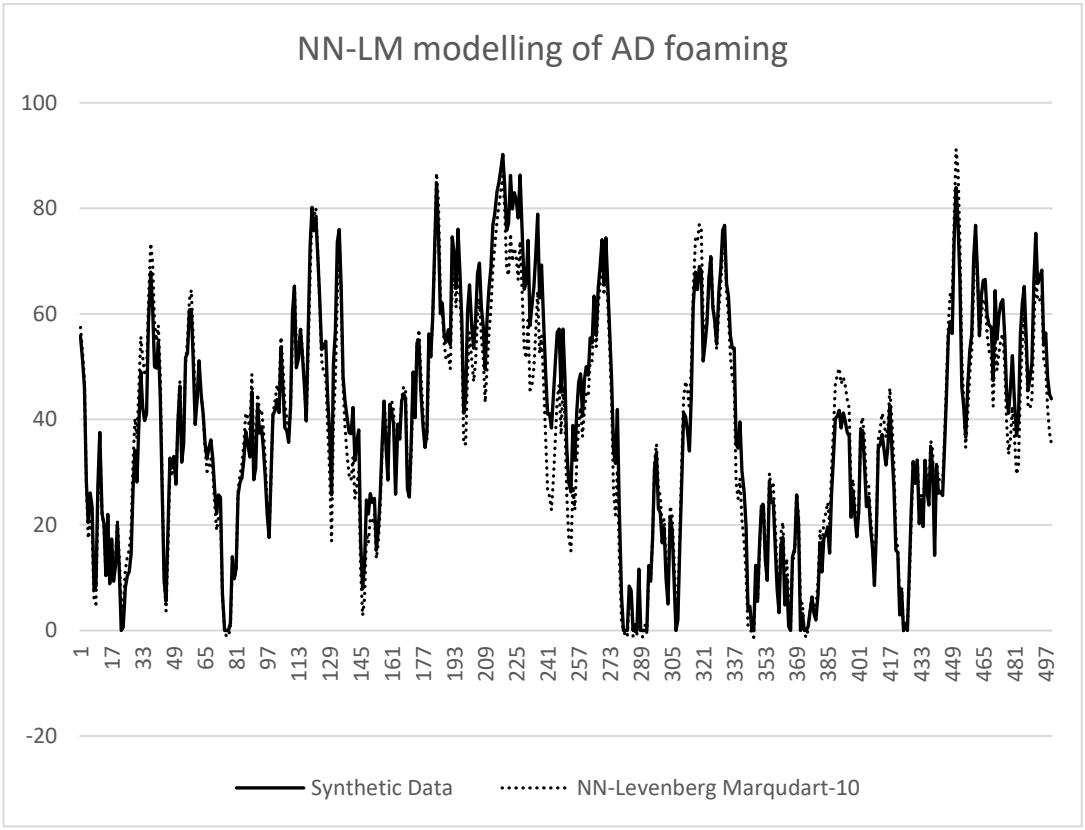


Figure 7-13: NN-Levenberg Marqudart-10 model

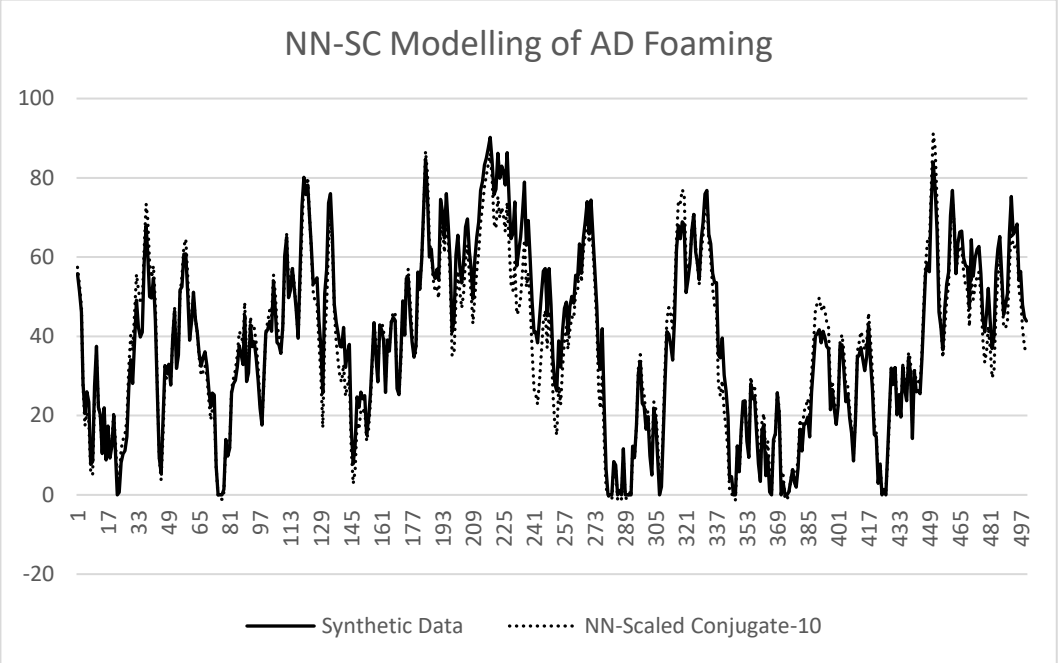


Figure 7-14:NN-Scaled Conjugate-10 Model validation

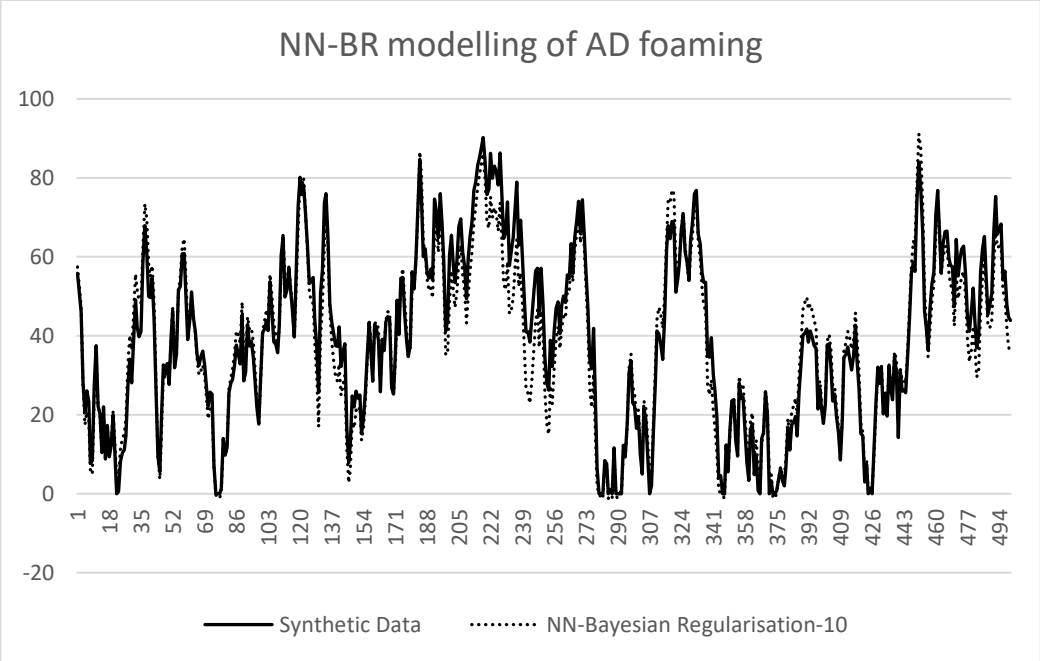


Figure 7-15:NN-Bayesian Regularisation-10 Validation model

7.3.2.1 Discussion on Neural Network models

In MATLAB NN fitting toolbox the default number of layers is 2, the default number of neurons in the hidden layer is 10, and the default training function

is NN-LM while the default transfer function for hidden layers is tansig and the default for the output layer is purelin.

Following the creation of the NN, the configure command is used to configure the network object and also initialize the weights and biases of the network in order to get the network ready for training as shown in figure 7-16

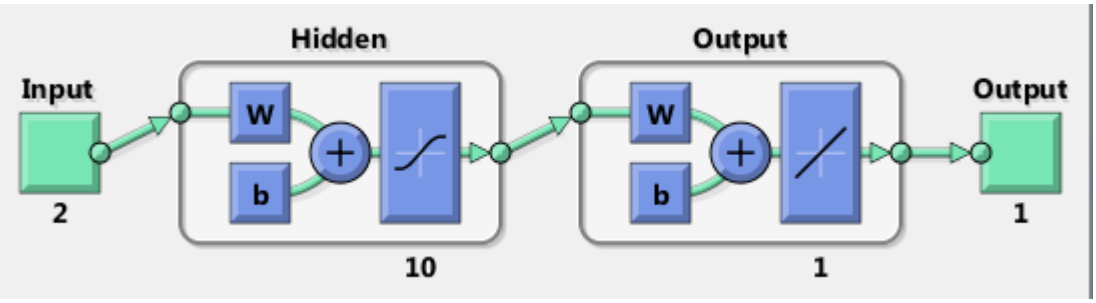


Figure 7-16: The structure of the neural network as configured in Matlab

The neural network training involved tuning the values of the weights and biases of the network to optimize network performance defined by the network performance function; mean square error (MSE). The batch mode of training NN was applied instead of the incremental mode as discussed in section 2.8.2.4.1. This is because batch training is significantly faster and produces smaller errors than incremental training as all the inputs in the training set are applied to the network before the weights are updated (Beale, et al., 2015).

As observed in this study, training multilayer feedforward networks can be achieved using any standard numerical optimization algorithm to optimize the performance function. The key difference between these optimization methods is that they use either the gradient of the network performance with respect to the network weights, or the Jacobian of the network errors with respect to the weights (Beale, et al., 2015). However, both are calculated using a technique called the backpropagation algorithm, which involves performing computations backward through the network that is derived using the chain rule of calculus (Beale, et al., 2015).

As the training progresses, theperformance is monitored and a performance plot plotted at the end of the network training as shown in figure 7-17.

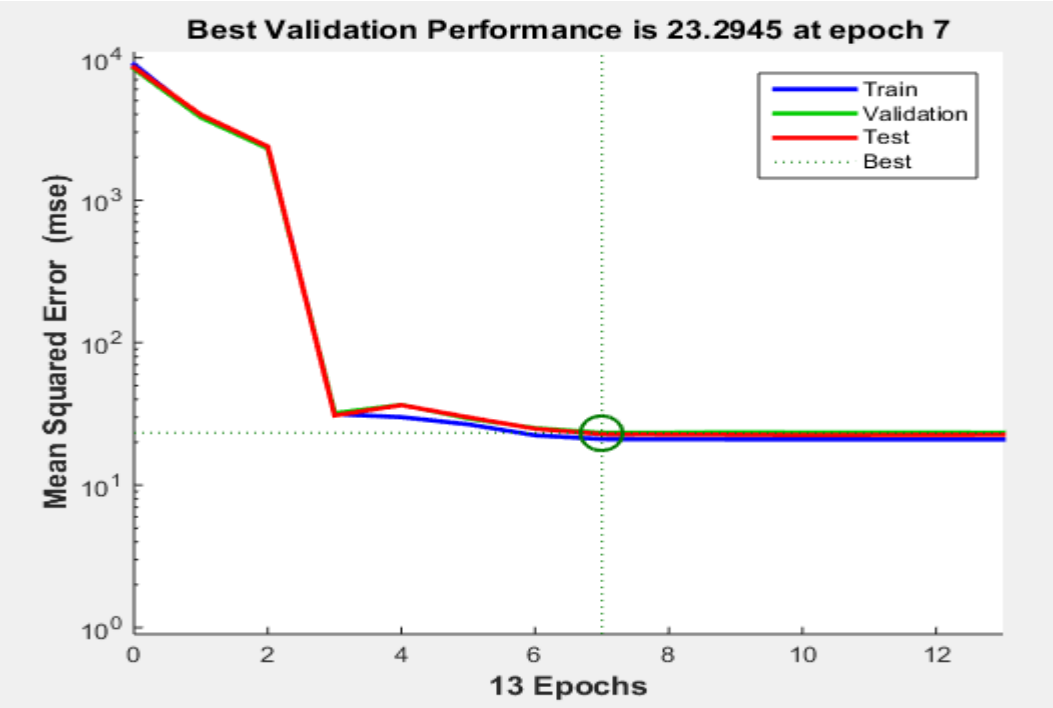


Figure 7-17: Neural network performance plot-NN-LM-10

The point with a green circle is the best performance during training as it incdicates the iteration at which the validation performance reached a minimum. However, the training continued for 6 more iterations before it stopped in order to assess over-fitting.

Hence, figure 7-17 does not indicate any problem during training as the validation and test curve is similar. However, if the test curve did increase significantly before the vlaidation curve increased then there is potential that overfitting took place.

The network is then validated using a regression plot that shows the relationship between the output of the network and the measured output as shown in figure 7-18.

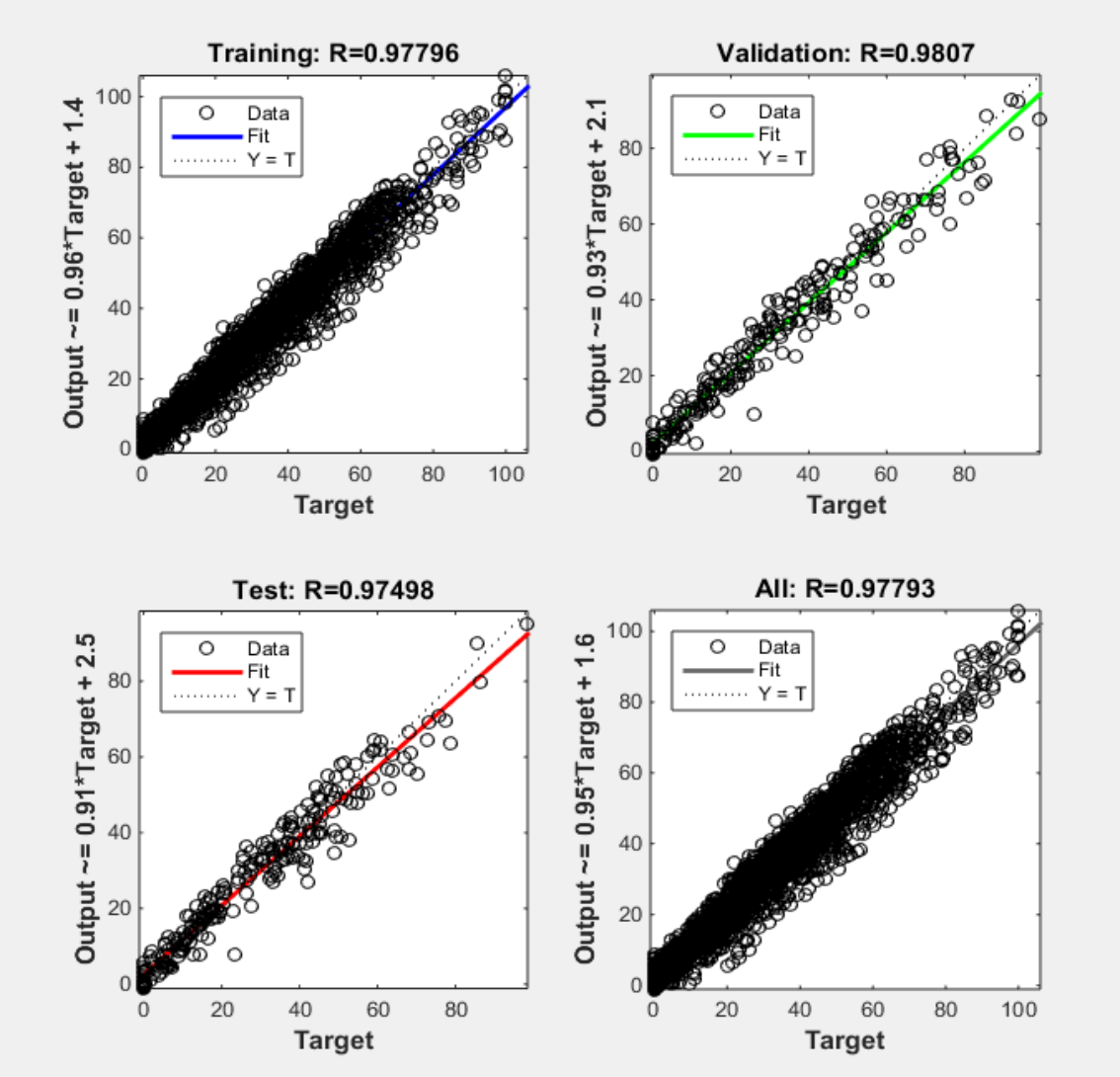


Figure 7-18: Regression plot of neural network NN-LM-10

The NN models were able to demonstrate that we could efficiently model foaming in AD hence provide the opportunity for automated control of the process.

Although the different NN algorithms performed alike during the final validation, it could be observed that during training, NN-Levenberg Marquardt performed better than the other algorithm for the different number of neurones trialled.

This conforms to the statement from Beale et al. (2015) that NN-LM is the fastest training function for feedforwardnet as it performs better on function fitting (nonlinear regression) problems than on pattern recognition problems. However, it tend to be less efficient for large networks (with thousands of

weights), since it will require more memory and more computation time for these cases. However, with the advancement in technology, this is not actually a challenge.

7.4 Conclusion

Based on the preceding discussion on modelling foaming in anaerobic digester using MFL, NNFIS and NN models, it is established from this research that NNFIS and NN models were more efficient at modelling the dynamics of foaming in AD.

The MFL model is highly dependent on a good knowledge of the system to be modelled. Hence, any feature of the process being modelled that was not captured during the creation of the MFL will not be reflected in the model validation. This leads to inadequate performance of the model.

On the other hand, NN model require huge amount of data to ensure the system has a good understanding of the system operating conditions to be able to make a good prediction. Thus, there is need for the model to be exposed to sufficient data that represent the potential characteristics of the process being modelled. This is crucial to achieving a high performance from the NN model.

Considering that there has been absence of a robust intelligent model for predicting foaming in AD, this study had adopted a systematic approach by establishing a thorough understanding of AD foam formation to create the best model for predicting foaming in AD. Thus, utilising intelligent model such as the NFIS or NN-LM model as developed in this research will be an invaluable asset that maximise efficiency and curb the occurrence of nuisance foaming in AD thereby accelerating wider adoption of AD

Chapter 8 : Conclusions and recommendations

This chapter presents a succinct summary of diverse research deductions that emanated from this study, limitations, and implication for future research in this area.

8.1 Study overview

Prior to summarising the research conclusion, it is germane to assess the extent to which the aim and objectives of this study were attained.

As a reminder, the study objectives as presented in chapter 1 are to:

1. Extensively review the literature on AD operation to ascertain factors relating to AD foaming and establish knowledge gaps that will assist in developing the research methodology
2. Comprehensively review the literature on modelling AD operation and foaming
3. Identify suitable full-scale AD in Scotland and collect data on operation parameters including foaming history and biogas production inventory
4. Carry out statistical analysis including correlation analysis to identify statistically significant process and design variables affecting foam formation.
5. Design and carry out laboratory experiment that assesses the variation in foam formation to variations in the identified significant process variables.
6. Carry out metagenomics and metabolomics analysis of foaming and non-foaming samples to enhance knowledge on microbial composition and their influence on foaming.
7. Formulate, calibrate, and validate model for predicting the formation and extent of foaming in AD with a view to improving the efficiency of biogas from such plants.

The first objective was to extensively review the literature on AD operation and modelling to ascertain factors relating to AD foaming and limitations in existing models in order to establish knowledge gaps that will assist in developing the research methodology. This objective was met in chapter 2. It was identified that although there have been many research efforts exploring

the cause of foaming in AD, none of them have applied a systematic statistical approach in underpinning and distinguishing foam causal factors from those factors that promote foaming in AD.

'Omics' studies were identified as a very useful high throughput tool that is vital for developing/validating comprehensive knowledge of the microbial and metabolic constituent of AD content; however, the high cost of such analysis limit its use for daily monitoring of the digestion process.

Another fact that was revealed by the literature search is that the most extensive and widely used AD mechanistic model (ADMI) did not model foaming occurrence as it is difficult to establish the kinetics of foaming occurrence. Thus, there exist one empirical model of risk of foaming in AD based on heuristic knowledge as record of foaming occurrence was not available.

Based on the literature review, the need arose to proactively mitigate the occurrence of foaming by developing predictive models (data driven) that could accurately ascertain the propensity for foaming and form the basis of a potential automated control mechanism.

Consequently, the following were identified as crucial for using data driven model to predict foaming occurrence in AD:

1. The need to select the most suitable variable,
2. Acquire data through properly designed laboratory experiments,
3. Apply synthetic data generation on collected data from laboratory experiment to extend data and make it sufficient for the data driven modelling approach.

The second objective was achieved in chapter four. This involved the identification of suitable full-scale AD in Scotland and collection of data on operating parameters including foaming history and biogas production inventory. Statistical analyses including correlation analyses were carried out to identify statistically significant process and design variables affecting foam formation.

In chapter five, the third objective was met as laboratory experiments were executed to assess the variation in foam formation to variations in the identified significant process variables identified in chapter four

By carrying out metagenomics and metabolomics analysis of foaming and non-foaming samples to enhance knowledge on microbial composition and their influence on foaming, the fourth objective was achieved which provided further evidence on the microbial population and metabolites in foaming and non-foaming digesters.

Objective five was accomplished in chapter 7 by formulating, calibrating, and validating model for predicting the formation and extent of foaming in AD with a view to improving the efficiency of biogas from such plants.

Hence, it could be observed that the objectives set for this study have been fully achieved.

8.2 Conclusions

Based on the study, the following conclusions have emerged.

8.2.1 This study focused on unravelling challenges militating against a clear understanding of factors causing foaming to separate it from factors that promote foaming occurrence. Applying a systematic approach of historical data analysis supported by laboratory experiment, it was identified that the most common cause of digester foaming is organic overload. Although other factors such as inefficient mixing and temperature variation were also identified as potential causes of foaming occurrence, it was discovered through survey carried out with Wastewater Treatment Plants (WWTP) operating AD that such factors as temperature and mixing are easily identified and addressed through the programmable logic control (PLC) system used currently in AD process.

8.2.2 In the full-scale AD's surveyed, feeding was controlled by high and low-level indicators in the sludge holding tank, thus the digesters were fed based on the volume of available sludge in the sludge holding tank. As a reactive approach, the samples of feed sludge were taken intermittently and analysed for volatile solid concentration which took approximately a day to

get the result, the dry solid concentration obtained through the process is then compared with the online dry solid measurement.

To check on the efficiency of the reactors, volatile solids concentration and organic loading rate are determined based on the offline measurements. In addition, the digestate samples are also taken intermittently and offline measurements of pH, VFA, Alkalinity, volatile solid reduction, and biogas compositions are also determined. However, these values do not reflect the exact condition existing in the feed nor in the digester content at that point of analyses leading to system perturbation before the necessary indicators could be identified. Hence the need to develop more proactive approach to curbing system perturbation using predictive models as achieved in this study.

8.2.3 Although modelling has proved to be a very useful tool in optimising various biological processes such as AD there was only one model incorporating the risk of AD foaming occurrence based on heuristic knowledge. Thus, the need to investigate other modelling paradigms for predicting foaming occurrence in AD. The decision to use data driven modelling approach to enable AD operators easily adopt the model requires extensive experiment to isolate the causal factors of AD foaming. The experiments proved that the OLR and VFA are the dominant variables in AD foaming occurrence.

8.2.4 It was discovered however, that increasing the organic loading rate beyond 2.5 kgVS/m^3 resulted to an increased production of more VFAs than can be converted to methane as the acid formers which release carbon dioxide during the acidogenesis stage proceed more quickly than the methane-forming microorganisms.

Furthermore, as more carbon dioxide is produced, it solubilises in the sludge to form carbonic acid which further increases the overall volatile acid content in the digester, lowering the pH of the digester and causing system imbalance. This leads to the digester becoming more acidic inhibiting the methanogens thereby limiting methane production.

Thus, with increased accumulation of VFA in the digester which also increases the surface tension, most of the methane produced are encapsulated

in the solution leading to digester foaming as methane has a very low solubility compared to carbon dioxide.

8.2.5 The metagenomics analysis identified that for foaming digesters, the acidogenic bacteria were more prevalent than in non-foaming digesters. Accessing the metabolomics result, it was identified that the prevailing biochemical pathway was for acidogenesis. This is related to reduction in the optimal functioning of the hydrogenotrophic methane that are usual capable of reducing the carbon dioxide content in the digester. This condition lead to increasing carbon dioxide presence with a decrease in methane content of the biogas.

8.2.6 The metagenomics analyses also showed that the phylum bacteroidetes and proteobacteria were found to be predominant with a higher relative abundance of 30% and 29% respectively while the phylum actinobacteria representing the most prominent filamentous foam causing bacteria such as *Norcadia amarae* and *Microthrix Parvicella* had a very low and cosistent relative abundance of 0.9% indicating that the foaming occurrence in the AD studied was not triggered by the presence of filamentous bacteria.

8.2.7 The result of the metabolomics analyses clearly illustrates that most of the factors that have been related to foaming in the past are the by-product of organic overload. However, the variable that was identified to respond highly to system perturbation during the experiment was VFA. Thus, the foam predicting model was developed using OLR and VFA of the feed as the input factors while maximum volume of foam as a percentage of the digester volume was used as the output variable thereby permitting ease of adoption in practice.

8.2.8 Initial exploratory statistical analyses of the data collected in the laboratory on tVFA in the feed and digestate, FI in the feed and digestate, percentage of carbon dioxide in the biogas, biogas production rate, pH, soluble protein in the digestate, alkalinity in the feed and digestate showed that most of the variables followed the normal distribution. Hence, although the number of data collected was relatively low due to the nature of the experiment and variable monitored (foaming), this unique attribute made the extension of the data using synthetic/stochastic data generation techniques plausible.

8.2.9 Thus, it was possible to adapt Markov AR (1) model, synthetic data generation model commonly used in hydrological studies to synthesize additional data for the calibration and validation of the data driven foaming prediction models. Extensive testing of the stochastic model showed that it reasonably preserved the mean, variance, and serial correlation of the measured data.

8.2.10 To ensure a robust selection of the most relevant model, different artificial intelligence models were deployed using the MATLAB platform; Mamdani Fuzzy logic, NNFIS and a variation of the neural network model (Bayesian regularisation, Levenberg-Marquardt and Scaled Conjugate). The NNFIS (Neural network fuzzy inference system) and NN (neural network-Levenberg Marqudart, Bayesian and Scaled conjugate) models performed better than the MFL(Mamdani Fuzzy logic model). The best performing model was NN-Levenberg Marqudart model, with RMSE = 5.49, MSE = 30.19 and R^2 = 0.9435.

8.3 Recommendations

8.3.1 As at the time of this research there was no commercially available on-line sensor for measuring VS and VFA. However, recently, advancement in technology has led to development of on-line sensors for measuring VFA (OPTI-VFA and Voice200Ultra SIFT-MS) but not VS. Thus, the development of affordable and commercially available on-line sensor for monitoring VS and VFA is crucial to wider implementation of AD in different sectors not just wastewater treatment.

8.3.2 There is need to carry out further experimental analysis using online sensor for measuring VS and VFA to verify the suitability of the NN-Levenberg Maqudart model in laboratory scale and full scale anaerobic digester.

8.3.3 This study has been limited to one stage digester; however, research has shown that two stage digesters could be more efficient. Hence it will be great if the test of the model using online sensors will also consider two stage digestion processes to evaluate the efficiency of the model.

8.3.4 The result of omics analysis created a more in-depth information on the functionality and adaptation of microbial population present in the anaerobic digester. For example, it was observed that Fibrobacter exhibited an extreme performance in the foaming and non-foaming digesters. Existing research studies highlighted its ability to effectively degrade cellulosic material, thus, it will be useful to explore how Fibrobacter could be effectively deployed in future biotechnology projects.

8.3.5 In addition, 3-Methyleneoxindole was identified as a key metabolite in the foaming AD thus will be useful to explore its application as a very useful biomarker in the development of biosensors for measuring foaming occurrence in AD.

References

IWA , 2002. *Anaerobic Digestion Model No.1 (ADM1).* s.l.:International Water Association.

Abbasian, F., Lockington, R., Megharaj, M. & Naidu, R., 2015. extraction and 2-DE for metaproteomics of miThe integration of sequencing and bioinformatics in metagenomics.. *Review in Environmental Science and Bio/Technology. 14(3), 357-383,* 14(3), pp. 357-383.

Adeloye, A. & De Munari, A., 2006. Artificial neural network based generalised storage-yield-reliability models using the Levenberg_Marquard algorithm. *Journal of Hydrology,* Volume 326, pp. 215-230.

Adeloye, A. J. & Rustum, R., 2012. *Water and Wastewater treatment (D20/1 WW).* Edinburgh: Heriot Watt University.

Aebersold, R., Teplow, D., Hood, L. E. & Kent, S., 1986. Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacyrlamide gels for direct sequence analys. *Journal Biol. Chem.,* Volume 261, pp. 4229-4238.

Ahring, B., Sandberg, M. & Angelidaki, I., 1995. Volatile fatty acids as indicators of process imbalance in anerobic digestors. *Applied Microbiology and Biotechnology,* Volume 43, pp. 559-565.

Alfons, A. et al., 2011. *Synthetic data generation of SILC Data.* Advanced Methodology for European Laeken Indicators ed. https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP6-D6.2-240611.pdf: European Commission.

Allen, S. D. & Isom, D. K., 2012. *Anaerobic Digester Design and Operation,* s.l.: United States Patenet.

American Public Health Association, 2014. *Standard methods for the examination of water and wastewater.* 22 ed. Washington: American Public Health Association.

Anderson, L. & Anderson, N. L., 1996. Twenty years of two dimensional electrophoresis: past, present and future. *PubMed,* Volume 17, pp. 443-453.

Angelidaki, I., Ellegaard, L. & Ahring, B., 1993. A mathematical model for dynamic simulation ofanaerobic digestion of complex substrates focusing on ammonia inhibition. *Biotechnology and Bioengineering,* Volume 42, pp. 159-166.

Antonopoulou, G., Alexandropoulou, M. & Lytras, C., 2015. Modeling of Anaerobic Digestion of Food Industry Wastes in Different Bioreactor Types. *Waste Biomass Valor,* 6(DOI: 10.1007/s12649-015-9362-7), p. 335.

APHA, 2012. *Standard Methods for the Examination of water and Wastewater.* 22 ed. Washinghton: American Public Health Association.

Arbib, M., 2003. *A handbook of brain theory and neural network.* 2nd ed. s.l.:MIT.

Arbib, M., 2003. *The handbook of brain theory and neural networks.* s.l.:MIT.

Arthur, S. & Pender, G., 2006. *Urban Drainage Design & Analysisr Scott Arthur, Professor Garry Pender.* Edinburgh: Heriot-Watt University.

Arumugam, T., Parthiban, L. & Rangasamy, P., 2015. Two phase anaerobic digestion model of a tannery solid waste: Experimental Investigation and modeling withn ANFIS. *Arab Journal of Science Enginering,* 40(DOI 10.1007/s13369-014-1408-9), pp. 279-288.

Astel, A., Tsakovski, S., Barbieri, P. & Simeonov, V., 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research,* 41(9), pp. 4566-4578.

Balk, M., Weijma, J. & Stams, A., 2002. Thermatoga lettingae sp. ; a novel thermophilic, methanol degrading bacterium isolated from a thermophilic anaerobic reactor.. *International Journal of Systematic and Evolutionary Microbiology,* Volume 52, pp. 1361-1368.

Balussou, D. et al., 2012. An Economic Analysis of Three Operational Co-digestion Biogas Plants in Germany. *Waste Biomass Valor,* pp. 3:23–41,DOI 10.1007/s12649-011-9094-2 .

Banfield, P., 2013. *Priciples of Wastewater Treatment and Waste to Energy.* Edinburgh: Velia Water Outsorcing LTD.

Banfield, P. & Littlejohn, J., 2013. *Waste 2 Energy_Seafield Case Study.* [Online] Available at: http://www.ice.org.uk/ice_web_portal/media/scotland/gwos-waste2energy-seafield-100113.pdf
[Accessed 24 February 2014].

Barber, W., 2014. Influence of wastewater treatment on sludge production and processing. *Water and Environment Journal,* 28(1), pp. 1-10.

Barjenbruch, M., H., H., O., K. & J., T., 2000. Minimising of foaming in digesters by pre treatment of the surplus sludge. *Water Science and technology,* 42(9), pp. 235-241.

Bartell, R., Matson, E., Mueller-Spitz, S. & Kleiheinz, G., 2015. Investigation of methanosarcinales and Methanomicrobiales presence within a dry anaerobic digester. *Journal of Microbiology Research,* 5(3), pp. 101-108.

Batstone, D. J., Jensen, P. & Ge, H., 2011. *Biochemical Treatment of Biosolids-Emerging Technologies.* Australia, Water, AWA publishing, pp. 90-93.

Batstone, D. J., Keller, J. & Steyer, J. P., 2006. A review of ADM1 extensions, applications and analysis: 2002 - 2005. *Water Science and Technology,* 54(4), pp. 1-10.

Batstone, D. et al., 2002. *Anaerobic Digestion Model No. 1,* London: IWA publishing.

Beale, D. et al., 2016. An 'omics' approach toward the characterisation of laboratory scale anaerobic digesters treating municipal sewage sludge. *Water Research,* Volume 88, pp. 346-357.

Beale, M., Hagan, M. & Demuth, H., 2015. *Neural network toolbox: User's guide,* Natick MA: MathWorks Inc.

Beale, M., Hagan, M. & Demuth, H., 2015. *Neural Network Toolbox_Users Guide,* s.l.: The Mathworks Inc..

Becker, S., 1991. Unsupervised learning procedures for neural networks. *International journal of Neural Systems,* Volume 7, pp. 17-33.

Ben-David, A. & Davidson, C., 2014. Estimation method for serial dilution experiments. *Journal of Microbiology Methods,* Volume 107, pp. 214-221.

Boe, K., 2006. *Online monitoring and control of the biogas process. PhD thesis, Institute of Environment and Resources,,* Denmark: Technical University of Denmark.

Boone, D. R., Whitman, W. & Rouviere, P., 1993. Diversity and taxanomy of methanogens. *In Methanogenesis,* Volume Springer, pp. 35-80.

Borell, G. et al., 2014. Comparative genomics highlightthe unique biology of Methanomassilicoccales, a Thermoplasmatales-related seventh order of methanogenic archea that encodes pyrrolysine. *BMC Genomics,* 15(679).

Bouallaguia, H., Touhamia, Y., Cheikhb, R. & Hamdia, M., 2005. Bioreactor performance in anaerobic digestion of fruit and vegetable wastes. *Process Biochemistry,* 40(http://dx.doi.org/10.1016/j.procbio.2004.03.007)), pp. 989-995.

Bradbury, S. & Evennett, P., 1996. *Fluorescence microscopy, contrast techniques in light microscopy.* Oxford, United Kingdom: BIOS Scientific Publishers, Ltd.

Buswell, A. M. & Mueller, H., 1952. mechanism of methane fermentation. *Journal of Ind Eng Chem,* 44(3), pp. 550-552.

Butler, D. & Davies, J. W., 2011. *Urban drainage.* 3rd ed. London: Spon.

Caparaso, J. G. et al., 2012. Ultra-high throughput microbial communityanalysis on the Illumina HiSeq and MiSeq Platforms. *ISME,* Volume 6, pp. 1621-1624.

Castillo, E. & Izard, J., 2015. Metagenomics for bacteriology. In: *Metagenomics for microbiology.* s.l.:s.n.

Cevallos-Cevallosa, J. M. et al., 2009. Metabolomic analysis in food sciences, a review. *Trends in Food Science & Technology ,* Volume 20, pp. 57-566.

Chao, Y., Guo, F., Fang, H. & Zhang, T., 2014. Hydrophobicity of diverse bacterial populations in activated sludge sludge and biofilm revealed by microbial adhesion to hydrocarbons assay and high-throughput sequencing. *Colloids and Surfaces B: Biointerfaces ,* Volume 114, pp. 379-385.

Chen, Y. & Hashimoto, A., 1980. Substrate utilisation kinetic model for biological treatment processes. *Biotechnology and Bioengineering,* Volume 22, p. 2081.

Chen, Y. R. & Hashimoto, A., 1979. *Kinetics of methane fermentation.* Proc. Symposium on Biotechnology in Energy Production and Conservation, John Wiley & SOns Inc., New York.

Chu, C. & Lee, D. J., 2004. *Chemical Engineering Science,* Volume 59.

Cinar, O., 2005. New tool for evaluation of performance of wastewater treatment plant: Artificial neural network. *Process Biochemistry,* Volume 40, pp. 2980 - 2984.

Cirne, D. et al., 2007. Anaerobic digestion of lipid-rich waste - Effects of lipid concentration. *Renewable Energy,* Volume 32, pp. 965 - 975.

Cotrell, J. & London, U., 1999. Probability-based pprotein identification by searching sequence databases using mass spectometry data. *Electrophoresis,* Volume 20, pp. 3551-3567.

Cottret L, M. P. A. V. M.-S. A. S. L. C. H., 2010. Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, Baumannia cicadellinicola and Sulcia muelleri, with their insect host, Homalodisca coagulata.. *PLoS Comput Biol ,* 6(9, pii: e1000904. doi: 10.1371/journal.pcbi.1000904.).

Cox, M., Cookson, W. & Moffat, M., 2013. Sequencing the human microbiome in health and diseases. *Human Molecular Genetics,* 22(1), pp. 88-94.

Craig, R. & Beavis, R., 2003. A method for redcuing the time required to match protein sequeces with thandem mass spectra. *Rapid Communication Mass Spectrom,* Volume 17, pp. 2310-2316.

Dalmaua, J. et al., 2008. *A neural network approach for selecting the most relevant variables for foaming in anaerobic digestion.* iEMSs, 4th Biennial Meeting of International Environmental Modelling and Software Society (iEMSs).

Dalmau, J. et al., 2009. *Validation of knowledge base risk model for biological foaming in anaerobic digestion simulation.* s.l., AIAI workshop proceeding.

DEFRA, 2002. *Sewage Treatment in the UK-UK Implementation of the EC Urban Waste Water Treatment Directive,* London, UK: DEFRA Publications.

Deng, R., 2002. *Information and Communications Security.* Proceedings of the 4th International Conference , Singapore.

Deng, Z., 2015. *Validation and improvement of the OPTI-VFA_MSc Thesis,* Netherlands: Delft University of Technology.

Dobell, C., 1932. *Anthony VAn Leeuwenhoek and his 'Little animals'; being some account of the author of protozoology and bacteriology and his multifarious discoveries in these disciplines..* acessed on 01/09/2016 ed. s.l.:https://archive.org/details/antonyvanleeuwen00dobe.

Donoso-Bravo, A. et al., 2011. Model Selection, identification and validation in anaerobic digestion: A review. *Water Research,* Volume 45, pp. 5347-5364.

Eastman, J. & Ferguson, J., 1981. Solubilisation of particulate organic carbon during the acid phase of anaerobic digestion. *Journal of Water Pollution Control Federation,* 53(3), pp. 352-366.

Eberhart, R. C. & Dobbins, R. W., 1990. *Neural Network PC Tools-A Practical guide.* 1 ed. San DIego, California: Academic Press Inc..

Eddy/AECOM, M. &., 2014. *Wastewater Engineering - Treatment and Resource recovery.* Fifth ed. New York: McGraw-Hill International .

Eder, B. & Schulz, H., 2006. *Biogas Praxis.* 3rd Edition, ISBN 3-936896-13-5, ed. Germany: ökobuch Verlag,.

Edman, P., 1949. A method for the determination of the amino acid sequence of peptides. *Arch. Biochem. Biophys.,* Volume 22, pp. 475-483.

Eikelboom, D., 2000. *Process Control of Activated Sludge Plants by Microscopic Investigation.* 1 ed. London: IWA Publishing.

Eng, J., McCormack, A. & Yates, J., 1994. An approachg to correlate tandem mass spectral data of peptide with amino acid sequences in a protein database. *J. AM Sco Mass Spectrometry,* Volume 5, pp. 976-989.

Eskicioglu, C. & Ghorbani, M., 2011. Effect of inoculum/substrate ratio on mesophilic anaerobic digestion of bioethanol plant whole stillage in batch mode. *Process Biochemistry,* Volume 46, pp. 1682-1687.

European Commission, 1998. Commision Directive 98/15/EC. *Official Journal of the European Communities,* 67(29).

Evans, T. & Hofmann, G., 2012. Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Evans, T.G. & Hofmann, G.E. (2012) Defining the limits of physiological plasticity: how gene expression can assess and prePhilosophical Transactions of the Royal Society B-Biological Sciences,* Volume 367, pp. 1733-1745.

Ferguson, R., Coulon, F. & Villa, V., 2016. Organic loading rate: A promising microbial management tool in anaerobic digestion. *Water Research ,* Volume 100, pp. 348-356.

Ferguson, R., Villa, R. & Coulon, F., 2014. Bioengineering options and strategies for the optimization of anaerobic digestion processes. *Environmental Technology Reviews,* 3(1, DOI: 10.1080/09593330.2014.907362), pp. 1-14.

Filbert, J. W., 1985. *A survey of anaerobic digester operations,* Boston: Specialty conference, Environmental Engineering, American Society of Civil Engineers.

Fischer, J., Iannotti, E. & Porter, J., 1984. Anaerobic digestion of swine manure at various influent solids concentrations. *Agricultural Wastes,* 3(11; https://doi.org/10.1016/0141-4607(84)90042-8), pp. 157-166.

Fritz, J. et al., 2013. From meta-omics to causality: experimental models for human microbiome research.. *Microbiome,* 1(4).

Fryer, M., O'Flaherty, E. & Gray, N. F., 2011. Evaluating the Measurement of Activated Sludge Foam Potential.. *Water* , 3(doi:10.3390/w3010424), pp. 424-444;.

Fry, J., 2000. Bacterial diversity and 'unculturables'. *Microbiology Today,* Volume 27, pp. 186-188.

Fukushima, A. et al., 2009. Integrated omics approaches in plant systems biology. *Current Opinion in Chemical Biology,* 13(5, doi.org/10.1016/j.cbpa.2009.09.022), pp. 532-538.

Ganidi, N., Tyrrel, S. & Cartmell, E., 2009. Anaerobic digester foaming causes- A review. *Bioresoursce Technology,* 100(23), pp. 5546 - 5554.

Ganidi, N., Tyrrel, S. & Cartmell, E., 2011. The effect of organic loading rate on foam initiation during mesophilic anaerobic digestion of municipal wastewater sludge. *Bioresource Technology 102 ,* Volume 102, p. 6637–6643.

Garcia-Ocho, F. et al., 1999. Kinetic model for anaerobic digestion of livestock manure. *Enzyme and Microbial Technology,* Volume 25, pp. 55-60.

Garza, D. R. & Dutilh, B. E., 2015. From culture to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cellular and Molecular Life Sciences,* 72(22), pp. 4287-4308.

Ghasemi, A. & Zahedias, S., 2012. Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocinol Metab,* 10(2; doi: 10.5812/ijem.3505), pp. 486-489.

Gracey, A. Y., 2007. Interpreting physiological responses to environmental change through gene expression profiling. *Journal of Experimental Biology ,* Volume 210, pp. 1584-1592.

Grady, C., Daigger, G. & Lim, H., 1999. *Biological Wastewater Treatment.* New York: Marcel Dekker Inc..

Graves, P. & Haystead, T., 2002. Molecular biolgists guide to proteomics. *Microbiology and Molecular Biology Reviews,* 66(1), pp. 39-63.

Guermazi, S. et al., 2008. Discovery and characterisation of a new candidate division by an anerobic sludge digester metagenomic approach. *Environmental Microbiology,* 10(8), pp. 2111-2123.

Guitian, J. & Joseph, D., 1998. How bubbly mixtures foam and foam control using a fluidized bed. *International Journal of Multiphase Flow,* pp. 1-16.

Gunnigle, E. et al., 2013. A functional approach to uncover the low-temperature adaptation strategies of the archaeon Methanosarcina barkeri. *Applied and Environmental Microbiology,* 79(14), pp. 4210 - 4219.

Gunningle, E. et al., 2015. Low-temperature anaerobic digestion is associated with differential methanogenic protein expression. *FEMS Microbiology Letters.,* 362(10).

Gupta, A. J. & Gupta, U., 2014. Next generation sequencing and its application. In: http://dx.doi.org/10.101016/B978-0-12-416002-6.00019-5, ed. *Animal biotechnology*. s.l.:Elsevier Inc..

Guyon, I. & Elisseef, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research,* Volume 3, pp. 1157-1182.

Habicht, K. et al., 2011. Comparative proteomics and activity of a green sulfur bacterium through the water column of lake Cadagno, Switzerland. *Environmental Microbiology,* Volume 13, p. 203215.

Hall, R., Brouwer, I. D. & Fitzgerald, M., 2008. Plant metabolomics and its potential application for human nutrition. *Physilogia Plantarum,* 132(2), pp. 162-175.

Handelsman, J. et al., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products.. *Chemistry and Biology,* 5(10), pp. 245-249.

Hanreich, A. et al., 2013. Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic fermentations indicate concerted plant biogass degradation. *Systematic and Applied Microbiology,* 36(5), pp. 330-338.

Hao, L. P., Lü, F., Shao, L. M. & He, P. J., 2013. Improving the performance of thermophilic anaerobic digester through recirculation of low hydrogen biogas. *Environmental Science and Health, Part A: Toxic/Hazardous Substances and Environmental Engineering,* pp. 48:11, 1431-1436, DOI: 10.1080/10934529.2013.781900.

Haykin, S., 2009. *Neural networks and learning machines / by Simon Haykin.*. 3 ed. s.l.:Pearson.

Heinken, A. et al., 2014. Functional metabolic map of Faecalibacterium prausnitzii, a beneficial human gut microbe. *Journal of Bacteriology,* Volume 18, pp. 3289-3302.

Hettich, R., Pan, C., Chourey, K. & Giannone, R., 2013. Metaproteomics: Harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal Chem.,* 85(9), pp. 4203-4214.

Hill, D. & Barth, C., 1977. A dynamical model for simulation of animal waste. *Journal of Water Pollution Control Association,* Volume 10, pp. 2129-2143.

Hill, D. & Nordstedt, R., 1980. Modelling techniques and computer simulation of agricultural waste treatment processes. *Agricultural Wastes,* Volume 2, pp. 135 - 156.

Hill, D. T., 1982. A comprehensive dynamic model for animal waste methanogenesis. *Transaction of the ASAE,* 25(5), pp. 1374-1380.

Hill, D. T., 1983. Simplified Monod Kinetics of Methane Fermentation of Animal Wastes. *Agricultural Wastes,* Volume 2, pp. 1-16.

Hipel, K. & Mcleod, A., 1994. *Time series modelling of water resources and environmental systems.* Amsterdam: Elsevier Sciences B.V..

Hirakata, Y. et al., 2015. Identification and detection of prokaryotic symbionts in the ciliate metopus from anaerobic granular sludge. *Microbes and Environment,* 30(4), pp. 335-338.

Hoag, J., 2008. *Synthetic data generation:theory, technique and applications,* University of Arkansas, Fayetteville, AR, USA: Doctoral Dissertation_ISBN: 978-0-549-66137-5.

Holubar, P. et al., 2003. Advanced controlling of anaerobic digestion by means of hierarchical neural networks. *Commun Agric Appl Biol Sci.*, 68(2), pp. 167-170.

Hori, T. et al., 2006. Dynamic transition of a methanogenic population in response to the concentration of volatile fatty acids in a thermophilic anaerobic digester. *Applied and Environmental Microbiology,* Volume 72, pp. 1623-1630.

Hotelling, H., 1933. Analysis of complex statistical variables into principal components. *Journal of educational psychology ,* 24(6), p. 417.

Houtmeyers, J., Eynde, E., Poffé, R. & Verachtert, H., 1980. Relations between substrate feeding pattern and development of filamentous bacteria in activated sludge processes. *European journal of applied microbiology and biotechnology,* 9(1), pp. 63-77.

Husain, A., 1998. Mathematical models of the kinetics of anaerobic digestion-A selected review. *Biomass and Bioenergy,* 14(5/6), pp. 561-571.

IWA Task Group for ADM 1, 2002. *Anaerobic Digestion Model No.1 (ADM1).* s.l.:International Water Association.

J Horiuchi, S. K. M. K. T. K., 2001. Modeling of pH response in continuous anaerobic acidogenesis by anartificial neural network. *Biochemical Engineering Journal,* Volume 9, pp. 199-204.

Jaenicke, S. et al., 2011. Comparative and joint analysis of two metagenomics datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS One,* Volume 6, pp. 1-5.

Jain, S., A.K., L., S.K., B. & Kudchadker, A. P., 1992. Modelling of hydrolysis controlled anaerobic digestion. *Journal of Chemical Technology and Biotechnology,* 53(4), pp. 337-344.

Jimenez, J. et al., 2015. Instrumentation and Control of anaerobic digestion process: a review and some research challenges. *Rev Environ Sci Biotechnology,* 14(DOI 10.1007/s11157-015-9382-6), pp. 615-648.

Kalteh, A., Hjorth, P. & Berndtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software,* 23(7), pp. 835-845.

Kamarthi, S. V. & S., P., 1999. Accelerating neural network training using weight extrapolations. *Neural Networks,* 12(9), pp. 1282-1299.

Kana, E., Oloke, J. K., Lateef, A. & Adesiyan, M., 2012. Modeling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. *Renewable Energy,* Volume 46, pp. 276-281.

Kanu, I., Aspray, T. & A., A., 2015a. Understanding and predicting foam in anaerobic digester. *International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering,* 9(10), pp. 1056-1060.

Kanu, I., Aspray, T. J., Scott, A. & Adeloye, A., 2015. *Laboratory experiment to select variables for predicting foaming in anaerobic digester.* Edinburgh, In proceedings of the Infrastructure and Environment Scotland Postgraduate Conference, Heriot Watt universit.

Karničar, K. et al., 2016. Integrated omics approaches provide strategies for rapid erythromycin yield increase in Saccharopolyspora erythraea. *Microbial Cell Factories,* 15(93, DOI: 10.1186/s12934-016-0496-5).

Kester, G., Ahuja, K., Zakovec, J. & Sparks, C., 2014. *Biogas to Transportation Fuels at WRRFs- Potential, Incentives, and Issues.* California: WERF Webcast.

Khalid, A. et al., 2011. The anaerobic digestion of solid organic waste. *Waste Management,* Volume 31, pp. 1737-1744.

Khan, M. E., 2012. A comparative study of white box, black box and grey box tesing techniques. *International Journal of Advanced Computer Science and Applications,* 3(6).

Kim, H., 2013. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod,* 38(1; 10.5395/rde.2013.38.1.52), pp. 52-54.

Kim, K. et al., 2009. Urine metabolomics analysis for kidney cancer detection and biomaker discovery. *Mol. Cell. Proteomics,* Volume 8, pp. 558-570.

Klose, J., 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humagenetik,* Volume 26, pp. 231-243.

Knowles, S., n.d. *Design and theory of basic Anaerobic Digester,* s.l.: s.n.

Kobayashi, Y., Shinkai, T. & S., K., 2008. Ecological and physiological characterization shows that Fibrobacter succinogenes is important in rumen fiber digestion - review.. *Folia Micobiol,* 53(3, doi: 10.1007/s12223-008-0024-z), pp. 195-200.

Kohonen, T., 1982. Self Organised formation of topological correct feature maps. *Biological Cybernetics,* Volume 43, pp. 59-63.

Konwar, K., Hanson, N., Pagé, A. & S.J., H., 2013. K MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics,* 14(202, doi: 10.1186/1471-2105-14-202).

Kushalappa, A. C., Vikram, A. & Raghavan, G., 2008. Metabolomics of headspace gasfor diagnosing diseases of fruit and vegetables after harvest. *Stewart postharvest Review,* article 10(4(2)).

Kythreotou, N., Florides, G. & Tassou, S. A., 2014. A review of simple to scientific models for anaerobic digestion. *Renewable Energy,* Volume 71, pp. 701- 714.

Layden, N. M., 2007. An evaluation of autothermal thermophilic aerobic digestion (ATAD) of municipal sludge in Ireland. *Journal of Environmental Engineering Science.*

Le Gall, G. et al., 2011. Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis extracts and irritable bowel syndrome. *Journal of Proteome Research. ,* Volume 9, pp. 4208-4218.

Lee, J. et al., 2014. Temporal variation in methanogen communities fo four different full scale anaerobic digesters treating food waste-recycling wastewater. *Bioresource Technology,* Volume 168, pp. 59-63.

Leimena, M. M. et al., 2013. A comprehensive metatranscriptome analysis pipeline and its validationusing human small intestine microbiota datasets. *BMC Genomics.,* 14(DOI: 10.1186/1471-2164-14-530), p. 530.

Leven, L., Nyberg, K. & Schnurer, A., 2012. Conversion of phenols anaerobic digestion of organic waste - Areview of important microorgnisms and impact of temperature.. *Journal of Environmental managemet,* Volume 95, pp. 99-103.

Li, L. et al., 2015a. Dynamics of microbial community in a mesophilic anaerobic digester treating food waste: Relationship between community structure and process stability.. *Bioresource Technology,* Volume 189, pp. 113-120.

Li, P., Champane, B. & Anderson, C., 2013. Biogas production performance of mesophilic and thermophilic anaerobic co-digestion with fat, oil and grease in semi continous flow digester: effect of temperature, hydraulic retention time and organic loading rate. *Environmental Technology,* pp. 34:13-14, 2125-2133, DOI: 10.1080/09593330.2013.824010.

Liu, Q. & Kim, S., 2008. Evaluation of membrane fouling models based on bench scale experiments: A comparison between constant flowrate blocking laws and artificial neural network (ANNs). *Journal of Membrane Science,* Volume 310, pp. 393-401.

Luo, G., Fotidis, I. & Angelidaki, I., 2016. Comparative analysis of taxanomic, functional, and metabolic patterns of microbiomes from 14 full scale biogas reactors by metagenomic sequencing and radiosotopic analysis. *Biotechnology for biofuels,* 9(51,DOI 10.1186/s13068-016-0465-6).

Lutz, S. A. A., Field, K. & Benning, L., 2015. Integrated 'Omics' , targeted metabolite and single-cell analyses of arctic snow algae functionality and adaptability. *Frontiers in Microbiology,* 6(1323, doi: 10.3389/fmicb.2015.01323).

Lyberatos, G. & Skiadas, I., 1999. Modelling of anaerobic digestion - A review. *Global Nest International Journal,* 1(2), pp. 63-76.

Maier, R., 2009. Bacterial Growth. In: *Environmental Microbiology.* s.l.: Academic Press. Inc. , pp. 37-54.

Mamdani, E., 1977. Application of fuzzy logic to approximate reasoning using linguistic systems. *Fuzzy Sets Systems,* Volume 26, pp. 1182-1191.

Mamdani, E. H. & Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *Int . Journal of Man Machine Studies,* 7(1), pp. 1-13.

Massart, N., Bates, R. & Neun, G., 2006. *Proceedings of the Water Environmental Federation-Residual Biosolids Management.* (19) 1099-1117; https://doi.org/10.2175/193864706783797078, Water Environment Federation.

Mathworks, 2015. *Fuzzy logic toolbox-Users guide,* Natick, MA: Mathworks Inc.

McGinnis, D., 2006. Fate of rising methane bubbles in stratified waters: How much methane reaches the atmosphere?. *Journal of Geophysical Research,* Volume 111.

McMahon, T. A. & Adeloye, A., 2005. *Water Resouurces Yield.* 1st ed. Denver, Colardo, USA: Water Resources Publications.

Mesquitaa, D. P., Amaral, A. L. & Ferreiraa, E. C., 2013. Activated sludge characterization through microscopy: A review on quantitative image analysis and chemometric methods. *Analytica Chimica Acta ,* Volume 802, p. 802 (2013) 14–28.

Mesquita, D., L.A., A. & Ferreira, E., 2013. Activated Sludge Characterisation through Microscopy: A Review on Quantative Image Analysis and Chemometric Techniques. *Analytica Chimica Acta,* Volume 802, pp. 14-28.

Miler Herman, 2003. Vaccum Retort Anaerobic Digestion (VRAD) system and processes. *US Patent 6632362.*

Mladenovska, Z., Dabrowski, S. & Ahring, B., 2003. Anaerobic digestion of manure and mixture of manure with lipids: biogas reactor performance and microbial community analysis. *Water Sience and Technology,* 48(6), pp. 271-278.

Moeller, L., Herbes, C., Müller, R. A. & Zehnsdorf, A., 2010. Formation and removal of foam in the process of Anaerobic Digester. *Landtechnik 65 ,* 3(Energy production), pp. 207-207.

Moestedt, J., J., M. & Nordell, E., 2015. Determination of Methane and Carbon Dioxide Formation Rate Constants for Semi-Continuously Fed Anaerobic Digesters. *Energies,* 8(doi:10.3390/en8010645), pp. 645-655.

Monod, J., 1949. The Growth of Bacterial Cultures. *Annual Review of Microbiology,* 3(371).

Montgomery, D., 1984. *Design and analysis of experiments.* Second ed. USA: John Wiley & Sons.

Montgomery, D. & Runger, G., 2011. *Applied statistics and probability for engineers.* Fifth ed. Arizona: John Wiley & Sons.

Moser, A., 1988. *Bioprocess Technolog: Kinetics and Reactors.* New york: Springer.

Moset, V. et al., 2015. Mesophilic versus thermophilic anaerobic digestion of cattle manure: Methane productivity and microbial ecology.. *Microbial Biotechnology,* 8(5), pp. 787-800.

Mosey, F., 1983. Mathematical modelling of the anaerobic digestion process: regulatory mechanisms for the formation of short-chain volatile acids from glucose. *Wat. Sci.Tech.,* Volume 15, pp. 209-232.

Mota, V., Santos, F., Araújo, T. & Amaral, M., 2015. Evaluation of titration methods for volatile fatty acids measurement: effect of the bicarbonate interference and feasibility for the monitoring of anaerobic reactors. *Water Practice & Technology,* 10(3), pp. 486-495; DOI: 10.2166/wpt.2015.056.

Motte, J. C. et al., 2013. Total solids content: A key parameter of metabolic pathways in dry anaerobic digestion .. *Biothechnology for biofuels,* 6(164).

Mou, X. et al., 2005. Flow-cytometric cell sorting and subsequent molecular analyses for culture-independent identification of bacterioplankton involved in dimethylsulfoniopropionate transformations. *Applied and Environmental Microbiology,* 71(3), pp. 1405-1416.

Muth, T. et al., 2013. Searching for a needle in a stack of needles: challenges in metaproteomics data analysis.. *Molecular Biosystems. 9, 578,* Volume 9, p. 578.

Narayanasamy, S., Muller, E., Sheik, A. & Wilmes, P., 2015. Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microbial Biotechnology,* 8(3, doi:10.1111/1751-7915.12255), pp. 363-368.

Ng, C. et al., 2010. Metaproteogenomic analysis of a dominant green sulphur bacterium from ACE lake, Antartica. *ISME Journal,* Volume 4, pp. 1002-1019.

Nguyen, T. & Walker, E., 2006. *A first course in fuzzy logic.* Third ed. s.l.:Chapman and Hall.

Nicoletta, K., Georgios, F. & Savvas, A. T., 2014. A review of simple to scientific models for anaerobic digestion. *Renewable Energy ,* Volume 71, pp. 701-714.

Niekerk, A. et al., 1987. Foaming in anaerobic digesters- a survey and laboratory investigatiion. *JOurnal of Water Pollution Control Federation,* 59(5), pp. 249-253.

Oerther, D., De Los Reyes, F., De Los Reyes, M. & Raskin, L., 2001. *Quantifying filamentous microorganisms in activated sludge before, during and after an incident of foaming by oligonucleotide probe hybridizations and antibody staining.,* s.l.: Water Research 35, 14, 3325-3336.

O'Farrell, P., 1975. High resolution two-dimensional electrophoresis of proteins. *Journal of Biol.Chem.,* Volume 250, pp. 4007-4021.

Olsson, G. et al., 2005. *Instrumentation, Control and Automation in Wastewater Systems.* ISBN 1900222833: IWA Publishing.

Olsson, J. et al., 2004. Neural Network for rainfall forecasting by atmospheric downscaling. *Journal of Hydrologic Engineering,* 9(1), pp. 1-12.

Ometto, F. et al., 2014. Impacts of microalgae pre-treatments for improvedanaerobic digestion: Thermal treatment, thermal hydrolysis, ultrasound and enzymatic hydrolysis. *Water Research,* Volume 65, pp. 350-360.

Oresic, M., 2009. Metabolomics, a novel tool for studies in nutrition, metabolism and lipid dysfunction.. *Nutrition, Metabolism, and Cardiovascular Diseases,* 9(11), pp. 816-824.

Osson, G. & Newell, B., 2001. *Wastewater Treatment Systems - Modelling, diagnosis and control.* London: IWA publishing.

Pitman, A., 1991. Design Considerations for Nutrient Removal Activated Sludge Plants. *Water Science and Technology,* 23(4-6), pp. 781-790.

Powell, W., 2007. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* ISBN 0470171553: Wiley.

Qdais, H., Hani, K. B. & Shatnawi, N., 2010. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm.. *Resources, Conservation and Recycling,* Volume 54, pp. 359-363.

Rabee, R. & Adebayo, A., 2009. *Modelling activated sludge wastewater treatment plants using artificial intelligence (fuzzy logic and neural network).* Edinburgh: Heriot Watt University.

Rabee, R. & Adeloye, A., 2012. Improved modelling of wastewater treatment primary clarifier using hybrid anns. *International journal of computer science and artificial intelligence,* 2(4), pp. 14-22.

Ransom-Jones, E., Jones, D. L., McCarthy, A. J. & McDonald, J. E., 2012. The Fibrobacteres: an important phylum of cellulose-degrading bacteria.. *Microb Ecol.,* 2(doi: 10.1007/s00248-011-9998-1), pp. 267-281.

Ransom-Jones, E., L., J. D. & Edwards, A. M. J., 2014. Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. *Syst Appl Microbiol,* 37(7; doi: 10.1016/j.syapm.2014.06.001), pp. 502-9.

Rappé, M. S. & Giovannoni, S. J., 2003. The uncultured microbial majority. *Annual Review of Microbiology,* Volume 57, pp. 369-394.

Raskin, L. et al., 1994. Quantification of methanogenic groups in anaerobic biological reactors by oligonucleotide probe hybridisation. *American Society of Microbiology,* 60(2), pp. 1241-1248.

Ringner, M., 2008. What is principal component analysis?. *Nat. Biotechnol.,* Volume 26, pp. 303-304.

Rocha, M., Cortez, P. & Neves, J., 2007. Evolution of neural networks for classification and regression. *Neurocomputing,* 70(16), pp. 2809-2816.

Rodriguez-Roda, I. et al., 2013. *Anaerobic Digester Foaming: occurrence and control in Spain,* s.l.: IWA; accessed on February,10, 2014. URL: http://www.redbiogas.cl/wordpress/wp-content/uploads/2013/07/IWA-11509.pdf.

Rosenberg, M., Gutnick, D. & Rosenberg, E., 1980. Adherence of bacteria to hydrocarbons: A simple method for measuring cell-surface hydrophobicity. *FEMS Microbiology Letters,* 9(1), pp. 29-33.

Ross, J., 2004. *Fuzzy logic with engineering applications.* Second ed. s.l.:Wiley.

Roume, H. et al., 2013. Biomolecular isolation framework for eco-systems biology. *ISME Journal,* Volume 7, doi: 10.1038/ismej.2012.72, pp. 110-121.

Rustum, R. & Adeloye, A., 2009. *Artificial Intelligence Modelling of Wastewater Treatment Plants: Theory, Applications and Limitations.* 1 ed. s.l.:VDM.

Rustum, R. & Adeloye, A., 2011. *Artificial Intelligence Modeling of Wastewater Treatment plants:Theory, applications and limitations.*. 1 ed. s.l.:VDM.

Rustum, R. & Adeloye, A., 2012. Improved modelling of wastewater treatment primary clarifier using hybrid anns. *International journal of computer science and artificial intelligence,* 2(4), pp. 14-22.

Rustum, R. & Adeloye, A., 2012. Improved modelling of Wstewater Treatment Primary Clarifier Using Hybrid Anns. *International Journal of Computer Science and Artificial Intelligence,* 2(4), pp. 14-22.

Ruzin, S., 2005. *Plant microtechnique and microscopy.* http://microscopy.berkeley.edu/Resources/book.html ed. Berkeley: University Press.

Salas, J. D., Deller, J., Yevjevich, V. & Lane, W., 1980. *Applied Modelling of Hydrologic Time Series.* Littleton CO: Water Resources Publications.

Sansone, F. & Martens, C., 1981. Determination of volatile fatty acids and turn over rate in organic rich marine sediments. *Marine Chemistry,* Volume 10, pp. 233-247.

Sanusi, N., Zaharim, A. & Sopian, K., 2013. *An initial review: Stochastic application in wind speed forecast.* Cambridge, USA, http://www.wseas.us/e-library/conferences/2013/CambridgeUSA/ENVMECH/ENVMECH-11.pdf.

Saye, R. I. & Sethian, J. A., 2013. Multiscale modeling of membrane rearrangement, drainage, and rupture in evolving foam. *Science (www.sciencemag.org),* Volume 340, pp. 720-724.

Scheele, G., 1975. Two dimensional gel analysis of soluble proteins. Characterisation of guinea pig exrcine pancreatic proteins. *Journal of Biol.Chem.,* Volume 250, pp. 5375-5385.

Schink, B., 1997. Energies of syntrophic cooperation in methanogenic degradation.. *Microbiology and Molecular Biology Reviews,* 61(2), pp. 262-280.

Schön, M., 2009. *Numerical modelling of anaerobic digestion process in agricultural biogas plants,* s.l.: EINGEREICHT AN DER LEOPOLD FRANZENS UNIVERSITÄT INNSBRUCK.

Sciuto, G. L., Susi, G., Cammarata, G. & Capizzi, G., 2016. *A spiking neural network-based model for anaerobic digestion process.* [Online]
Available at: http://ieeexplore.ieee.org/document/7526003
[Accessed 30 4 2018].

Scottish Water, 2016. *Scottish Water Annual Report,* Scotland: Scottish water.

Sezign, M., Jenkins, D. & S., P. D., 1978. A unified theory of filamentous activated sludge bulking. *WPCF,* Volume 50, pp. 362-381.

Shah, F. et al., 2014. Microbial ecology of anaerobic digesters: The key players of anaerobiosis. *The Scientific World Journal,* Volume 2014.

Shah, N., Tang, H., Doak, T. & Ye, Y., 2011. *Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics.* Pacific Symposium on Biocomputing 2011, Kohala Coast, Hawaii, USA, s.n.

Siebels, A. M., 2011. *Foaming phenomenon in bench scale Anaerobic Digesters,* s.l.: University of Wisconsin-Madison.

Siedlecka, E. et al., 2008. Determination of Volatile Fatty Acids in Environmental Aqueous SamplesJ. Gajdus, Z. Kaczyński, P. Stepnowski. *Polish J. of Environ. Stud. ,* 17(3), pp. 351-356.

Silvestre, G. et al., 2011. Biomass adaptation over anaerobic co-digestion of sewage sludge and trapped grease. *Bioresoure Technology,* Volume 102, pp. 6830-6836.

Someone, 2013. *Literature Review on cause and contributors to foaming in Anaerobic Digesters.* s.l., s.n.

Soundharajan, B., Adeloye, A. & Remesum, R., 2016. Evaluating the variability in surface water resevoir planning characteristics during climate change impacts assessment. *Journal of Hydrology,* 538(DOI: 10.1016/J.hdrology.2016.04.051), pp. 625-639.

Srivastava, V. & Isaacson, R., 1991. Biogasification kinetics of biomass/waste blends. *Methane from Community Wastes,* pp. 101-111.

Stewart, D. et al., 2007. Metabolomic aproach to identifying bioactive compounds in beries: advances toward fruit nutritional enhancement. *Molecular Nutrition & Food Research ,* 51(6), pp. 645-651.

Steyer, J., Buffiere, P., Rolland, D. & Moletta, R., 1999. Advanced control of anaerobic digestion process through disturbance monitoring. *Water Res.,* 33(9), pp. 2059-2068.

St-Pierre, B. & Wright, A. D., 2014. Comparative metatgenomic analysis of bacterial populations in three full-scale mesophilic anaerobic manure digesters. *Applied Microbiology and Biotechnology,* 98(6), pp. 2709-2717.

Strauber, H., Schroder, M. & Kleinsteuber, S., 2012. Metabolic and microbial community dynamics during the hydrolytic and acidogenic fermentation in a leach-bed process. *Energy, Sustainability and Society.*, 2(1), pp. 1-10.

Strik, D. P. et al., 2005. Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB Neural Network Toolbox. *Environmental Modelling and Software,* , 20(6), pp. 803-810.

Subamanian, G., 2012. *A Literature review and survey of Anaerobic Digester foaming in waste water treatment plants.* s.l., s.n.

Subramanian, B. et al., 2012. *Survey of causes and prevention/control of anaerobic digester foaming- A WERF study*, s.l.: WERF.

Subramanian, B. & Pagilla, K. R., 2014. Anaerobic digester foaming in full-scale cylindrical digesters – Effects of organic loading rate, feed characteristics, and mixing. *Bioresource Technology*, Volume 159, pp. 182-192.

Tabb, D., McDonald, W. & Yates III, J., 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of Proteome Research*, pp. 21-26.

Tchobanoglous, G. & Burton, F., 1991. *Wastewater Engineering: Treatment, Disposal, and Reuse,*. 3rd ed. New York: Irwin McGraw Hill, Inc..

Tong, R. .., Beck, M. .. B. & Latten, A. .., 1980. Fuzzy Control of the Activated Sludge Wastewater Treatment Process. *Automatica ,* Volume 16, pp. 695-701.

Torkkola, K., 2003. Feature extraction by Non Parametric Mutual Information Maximization. *Journal of Machine Learning Research,* Volume 3, pp. 1415-1438.

Tuli, V. & Moyed, H. S., 1969. The Role of 3-Methyleneoxindole in Auxin Action. *The Journal Of Biological Chemistry*, Volume 244, pp. 4916-4920.

Tyers, M. & Mann, M., 2003. From genomics to proteomics.. *Nature,* Volume 422, pp. 193-197.

USEPA, 1989. *POTW Sludge Sampling and Analysis Guidance Document.* EPA 833-B-89-100 ed. USA: EPA Office of Water (4203).

Van Trees, H. L., 1968. *Detection, Estimation, and Modulation Theory.* Part 1 ed. New York: Wiley.

Vardar- Sukan, F., 1998. Foaming: consequences, prevention and destruction. *Biotechnology advances,* 16(5-6), pp. 913-948.

Varley, J. et al., 2004. Dynamic mulpti-point measurement of foam behaviour for a continuous fermentation over a range of key process variables. *Biochemical Engineering Journal*, Volume 20, pp. 61-72.

Vasconcelos de Sáa, L. et al., 2011. Simultaneous analysis of carbohydrates and volatile fatty acids by HPLC for monitoring fermentative biohydrogen production. *International*

Journal of Hydrogen Energy, 36(23; http://doi.org/10.1016/j.ijhydene.2011.08.056), p. 15177–15186.

Vavilin, V., Rytov, S. & Lokshina, L., 1996. A description of hydrolysis kinetics in anaerobic degradation of particulate organic matter. *Bioresource Technology,* Volume 56, pp. 229-237.

Vera, I., Sáez, K. & Vidal, G., 2013. Performance of 14 full-scale sewage treatment plants:Comparison between four aerobic technologies regarding effluent quality, sludge production and energy consumptionl. *EnvironmentalTechnology,* p. DOI:10.1080/09593330.2013.765921.

Villa, R. & Molinari, F., 2008. Reduction of Carbonylic and Carboxylic Groups by Plant Cell Cultures. *Journal of Natural Products,* 71(DOI:10.1021/np070386s), pp. 693-696.

Walker, M., Zhang, Y., Heaven, S. & Banks, C., 2009. Potential Errors in the Quantitative Evaluation of Biogas Production in Anaerobic Digestion Processes. *Bioresource technology,* 100(24), pp. 6339-6346..

Wang, D.Z., Xie, Z. & Zhang, S., 2014. Marine metaproteomics: current status and future directions. *Journal of proteomics,* Volume 97, pp. 27-35.

Wang, H. et al., 2010. Metaproteomic characterisation of a soil microbial community following carbon amnedment. *Soil Biol. Biochem,* Volume 42, pp. 48-56.

Water UK, 2017. *Water UK-Waste +Wastewater.* [Online]
Available at: http://www.water.org.uk/policy/environment/waste-and-wastewater
[Accessed 04 April 2017].

Weaire, D., Hutzler, S., Cox, S. & Kern, N., 2003. The fluid dynamics of foams.. *Journal of Physics: Condensed Matter,* Volume 15, pp. 65-73.

WEFTEC, 2006. *Design and operational considerations to avoid excessive AD foaming,* s.l.: s.n.

WEFTEC, 2007. Anaerobic Digestion. In: *Operation of Municipal Wastewater Treatment Plants.* s.l.:WWW. WEFNET.org, pp. 30-3 to 30-58.

WEFTEC, 2007. Anaerobic Digestion. In: www.wefnet.org, ed. *Operation of Municipal Wastewater Treatment Plants.* s.l.:WEFTEC, pp. 30-3 to 30-58.

Weinrich, S. & Nelles, M., 2015. Critical comparison of different model structures for the applied simulation of the anaerobic digestion of agricultural energy crops.. *Bioresource Technology,* Volume 178, pp. 306-312.

Westlund, A., Hagland, E. & M., R., 1998. Operational aspects on foaming in digesters caused by Microthrix parvecilla. *Water science and technology,* 38(8-9), pp. 29-34.

Wijekoon, K., Visvanathan, C. & Abeynayaka, A., 2011. Effect of organic loading rate on VFA production, organic matter removal and microbial activity of a two-stage thermophilic anaerobic membrane bioreactor. *Bioresource Technology,* 102(9), pp. 5353-5360.

WIlliams, T. O. & Tim, S., 2012. *Mixing Anaerobic Digester with high strenght waste addition.* Amherst, Massachusetts, North East Residuals & Biosolids Conference.

Wintermute, E. & Silver, P., 2010. Emergent cooperation in microbial metabolism. *Mol Syst Biol*, 6(407, doi: 10.1038/msb.2010.66).

Wishart, D., 2008. Application of metabolomics in drug discovery and development. *Drugs in R&D,* 9(5), pp. 307-322.

Wishart, D., 2008. Metabolomics: application to food science and nutrition research. *Trends in food science and technology,* Volume 4, pp. 482-493.

Woese, C. & Fox, G., 1977. Phylogenetic structure of the prokaryotic domain: tne primary kingdoms.. *PNAS,* 74(1), pp. 5088-5090.

Wolf, C., 2013. *Simulation, optimisation and instrumentation of agricultural biogas plants,* Maynooth: National University of Ireland.

WPCF, 1968. *Water Pollution Control Federation, Manual of Practce No 16 "Anaerobic Sludge Digestion",* Washington DC: WPCF.

WRAP ; NNFCC, 2012. *Anaerobic digestion infrastructure in the UK: September 2011.* UK: s.n.

WRAP, 2010. *Food Waste Chemical Analysis,* UK: WRAP.

WRAP, 2017. *Waste Resource Action Programme_Operational AD sites.* [Online]
Available at: http://www.wrap.org.uk/content/operational-ad-sites
[Accessed 24 March 2017].

WRAP & NNFCC, 2015. *The Official Information Portal on Anaerobic Digestion.* [Online]
Available at: http://www.biogas-info.co.uk/index.php/ad-map.html
[Accessed 28 January 2015].

Xie, R. et al., 2007. Full-scale demonstration of an ultrasonic disintegration technology in enhancing anaerobic digestion of mixed primary and thickened secondary sewage sludge. *Environ. Eng. Sci. ,* p. 6: 533–541 .

Y. Bajón Fernández, A. S. ,. R. V. ,. P. V. ,. E. C., 2014. Carbon capture and biogas enhancement by carbon dioxide enrichment of anaerobic digesters treating sewage sludge or food waste. *Bioresource Technology ,* Volume 159, pp. 1-7.

Yates, G. T. & Smotzer, T., 2007. On the lag phase and initial declineof microbial growth curves. *J. Theoret. Biol. ,* Volume 244, pp. 511-517.

Yin, R. K., 2014. *Case study research-Design and methods.* 4 ed. USA: SAGE publications.

Zadeh, L., 1965. Fuzzy Set. *Information and Control,* Volume 8, pp. 338-353.

Zadeh, L., 1999. From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems,* Volume 45, pp. 105-119.

Zakrzewski, M. et al., 2012. Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing.. *J Biotechnol. ,* 158(4; doi: 10.1016/j.jbiotec.2012.01.020.), pp. 248-258.

Zarraonaindia, I., Smith, D. P. & Gilbert, J. A., 2013. Beyond the genome:community-level analysis of the microbial world. *Biology & Philosophy,* Volume 28, pp. 261-282.

Zhang, A. et al., 2012. Modern analytical techniques in metabolomics analysis. *Analyst,* Volume 137, p. 293.

Zhang, B. M. K. et al., 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communication Mass Spectrometry,* Volume 17, pp. 2337-2342.

Zhang, W., Li, F. & Nie, F., 2010. Integrating multiple ‘omics’ analysis for microbial biology: application and methodologies. *Microbiology,* 156(DOI 10.1099/mic.0.034793-0), pp. 287-301.